现代数学丛书

陈希孺　赵林城　著

线性模型中的M方法

M-METHODS IN LINEAR MODEL

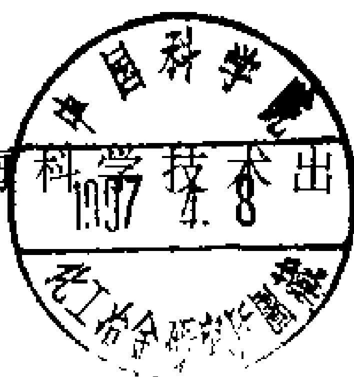
CHEN XIRU
ZHAO LINCHENG

上海科学技术出版社

·现代数学丛书·

线性模型中的 *M* 方法

陈希孺
赵林城 著



上海科学技术出版社

责任编辑 赵序明

·现代数学丛书·

线性模型中的 *M* 方法

陈希孺
赵林城 著

上海科学技术出版社出版、发行

(上海瑞金二路450号)

新华书店上海发行所经销 商务印书馆上海印刷厂印刷

开本 787×1092 小 1/16 印张 13.5 插页 4 字数 169,000

1996年10月第1版 1996年10月第1次印刷

印数 1-1,200

ISBN 7-5323-3907-6/O·192

定价：27.00 元

内 容 提 要

本书的内容是作者与合作者在一段研究工作成果的基础上，系统地论述线性回归模型的 M 方法的理论。其中包括：关于 M 估计定义的讨论；M 估计的强、弱相合性；M 估计的渐近正态性；基于 M 估计的线性假设检验的方法和理论；M 估计的线性表示等。

本书所反映的上述研究工作是国家自然科学基金资助的重点课题之一。

本书的读者对象：数理统计专业的研究工作者、实际工作者；大专院校概率统计专业的教师、研究生、高年级学生。

Modern Mathematics Series

M-METHODS IN LINEAR MODEL

Chen Xiru

Zhao Lincheng

Shanghai Scientific & Technical Puhlishers

M-METHODS IN LINEAR MODEL

Chen Xiru
Zhao Lincheng

Abstract

This monograph, based on the research achievements of the authors and their co-workers, gives a systematic account on the asymptotic theory of M-methods in linear regression models. If consists of a discussion on the definition of the M-estimate, on the weak and strong consistency and asymptotic normality, weak and strong linear representation of M-estimate, and linear hypothesis testing based on M-estimates. A series of special methods are developed in the book, enabling us to give very sharp results on a number of basic questions.

出版说明

从60年代起，由华罗庚教授任主编的《现代数学丛书》编辑委员会曾组织编著，并由我社出版了多部具有很高水平的数学学术专著，有几部专著并已在国外出了外文版，受到国内外数学界和广大读者的高度重视，获得了很高的评价。原编委会中华罗庚、关肇直、吴新谋三位教授虽已先后逝世，但他们为本《丛书》所作出的贡献迄今仍为人们所敬仰、怀念。由于某些客观原因，《现代数学丛书》的出版工作曾一度停顿。

为了适应现代数学的迅速发展，更好地反映我国数学家近几年的优秀研究成果，必须大力加强《现代数学丛书》的规划、编辑、出版工作，充实编委会的力量。考虑到不少编委年事已高，经向原编委会中大部分同志及数学界有关专家广泛征求意见后，于1990年对编委会作了调整，补充了一些著名的中年数学家和学科带头人，建立了新的编委会，并进一步明确了本丛书的宗旨。

《现代数学丛书》新的编辑委员会由苏步青教授任名誉主编、谷超豪教授任主编，18位著名数学家任委员。编委会负责推荐（或审定）选题和作者，主持书稿的审核等工作。

《现代数学丛书》的宗旨是：向国内外介绍我国比较成熟的、对学科发展方向有引导作用的、国内第一流水平的数学研究成果，反映我国数学研究的特色和优势，扩大我国数学研究成果的影响，促进学科的发展和国内外的学术交流。

为了实现上述宗旨，本丛书将陆续组织出版在基础数学、应用数学和计算数学方面处于学科发展前沿、有创见且具有系统完整

研究成果的现代数学学术专著。

为出版好《现代数学丛书》，我们热切地期望着数学界各位专家的大力支持和悉心指导，并欢迎广大读者提出宝贵的建议和意见。

上海科学技术出版社

序　言

理论研究和实践经验表明，线性回归分析中最常用的方法——最小二乘法，在一些情况下表现不理想。近几十年来，统计学家提出了许多替代方法供选择使用。本书的主题——M方法，就是其中之一，并且可以说是最受重视、研究成果最多的一种。

这个方法的研究发端于Huber在1964年的一项工作，其中考察了位置参数的M估计问题。于1973年，他把这个方法用于一般的多元线性模型。自那以后，这个领域受到统计学家相当的重视，20余年来积累了一批研究成果。

本书作者及其合作者在80年代后期开始涉足这个领域。所得到的研究成果就是写作本书的基础。也可以说，本书也就是作者及合作者这一段工作的总结。当然，工作还没有结束，有的问题还有待于在今后取得根本的进展。

本书内容包括：对M估计定义的讨论；M估计的强、弱相合性；渐近正态性；基于M估计的线性假设检验的方法和理论（这在某种意义上可看作是古典方差分析的一种推广）；以及M估计的线性表示，即用一个线性独立和去逼近M估计。我们对M估计的一个特例——最小一乘估计，给予了较多的注意。除了其在应用上的重要性外，这个特例对一般情况下的理论发展有着启示的作用。至于最小二乘估计，它当然无疑是M估计家族中最重要的成员。但本书对它着墨不多。因为关于这个题目前此已有专著[5]。

本书只限于考察线性回归且随机误差独立的情况。这个情况在统计上是最常用的。另外，这个情况的相对简单的架构，有利于

得到系统深刻而非表面的结果。当然,M 方法的使用不限于线性模型,例如,有一些学者已尝试将其用于带线性部分的半参数模型,并取得了一些成果。我们觉得,如果把线性情况研究透彻了,弄清了在这一情况下的一些主要问题能推进到何种程度,就可以树立一个标杆,作为在更复杂的模型下努力的目标。

在本书写作中,我们努力遵循深入浅出的原则,凡是能用初等工具处理的问题,就尽量采用初等工具。尽管本书中不少定理的证明很是曲折繁复,但所用的预备知识却不多,具有理工科概率统计和初步的分析及矩阵知识,即可顺利阅读本书。书中不加证明地引用了少量概率不等式和在一般教本中不易见到的极限理论,也都介绍了其出处,读者易于从国内常见的资料中查到。本书可以适合广大的读者,包括对线性统计模型大样本理论感兴趣的教师、研究人员和研究生、大学生;以及对 M 方法感兴趣的实用工作者等。

本书的主题是国家自然科学基金一项重点课题的一部分。在多年的工作进程中,我们一直得到该基金的资助,谨在此表示感谢;华东师范大学王静龙教授审阅了本书全稿,并指出和纠正了个别笔误之处,在此一并致谢。

由于作者水平所限,对书中一些不妥以至错误之处,希望广大读者和专家同行不吝指教。

作 者

1994 年 6 月于中国科学技术大学研究生院

目　录

CONTENTS

第 1 章

M 估计的定义

首先交代一下本书中常用的一些符号. 数、向量、矩阵和函数，如无特别申明，都是实的. 不加“′”的向量总是指列向量. 若 a 为向量或矩阵，则以 a' 记其转置，以 $\|a\|^2$ 记 a 的各元的平方和，以 $|a|$ 记 a 各元绝对值的最大值. 当 A 为方阵时，$A>0$、$A\geqslant 0$ 分别表示 A 为正定和非负定的. 当 A、B 为同阶方阵时，$A\geqslant B$ 表示 $A-B\geqslant 0$. 当不致引起混淆时，常把极限号下的“$n\to\infty$”等略去不书. p 维欧氏空间记为 R^p.

§1.1 用极值定义 M 估计

一、线性回归模型

本书常把线性回归模型写为下面的形式：

$$Y_i = x_i'\beta_0 + e_i \quad (1 \leqslant i \leqslant n, n \geqslant 1). \tag{1.1}$$

这里 β_0 是未知的 p 维回归参数向量；Y_i、x_i 和 e_i 分别是第 i 次观察或试验时的目标变量(因变量)、解释变量(自变量)和随机误差的取值. 注意，后者是不可观察的，因而其值未知；而 Y_i 和 x_i 已知，分别为 1 维和 p 维. 在多数情况下，x_i 是一个已知的、不带随机性的向量，在理论论证时，我们把它作数值向量处理. 本书个别地方涉及 x_i 本身也是某一个 p 维随机向量的观察值的情况，这时我们用大写的 X_i 来记，以示区别. 在理论论证时，X_i 要作为随机向量处理. 模型(1.1)的统计问题，就是要利用观察或试验所得的

数据(x_i,Y_i) $(1\leqslant i\leqslant n)$去对 β_0 作出推断(估计、检验等). 在有些问题中,也涉及对误差 e_i 的分布特征(如 e_i 的方差)的推断问题.

通常在应用中,线性回归函数常包含一常数加项,这当然是(1.1)的特例,只须取 x_i 的第一分量为 1 即可. 在有些情况下,有必要将这常数项明显标出,这时我们将模型写为:

$$Y_i = \alpha_0 + x_i'\beta_0 + e_i \quad (1 \leqslant i \leqslant n, n \geqslant 1), \tag{1.2}$$

x_i、β_0 仍为 p 维. 一个重要情况是(1.2)中只有常数项部分:

$$Y_i = \alpha_0 + e_i \quad (1 \leqslant i \leqslant n, n \geqslant 1), \tag{1.3}$$

(1.3)称为*位置参数模型*,α_0 称为*位置参数*.

还需要说明的一点是:参数 β_0 的取值范围称为*参数空间*. 如无特殊指定,就把这参数空间记为 R^p.

二、用极值点定义 β_0 的 M 估计

考虑线性模型(1.1),选定一个定义于 R^1 的函数 ρ,令

$$H_\rho(\beta) = \sum_{i=1}^{n}\rho(Y_i - x_i'\beta) \quad (\beta \in R^p). \tag{1.4}$$

β_0 的 M 估计记为 $\hat{\beta}_n$,定义为 $H_\rho(\beta)$的一个最小值点:

$$H_\rho(\hat{\beta}_n) = \min\{H_\rho(\beta) : \beta \in R^p\}. \tag{1.5}$$

注意:这里并未要求 $\hat{\beta}_n$ 有唯一性,而 $\hat{\beta}_n$ 也的确可以不唯一. 一个常提到的重要例子是对模型(1.3)若取 $\rho(u)=|u|$,则 $\hat{\beta}_n$ 就是 $Y_1,\cdots,Y_n$ 的样本中位数 $\mathrm{med}(Y_1,\cdots,Y_n)$,而当 n 为偶数时,后者一般不唯一. 对一般模型(1.1)也不难举出 $\hat{\beta}_n$ 不唯一的例子来.

这种估计是 Huber 最先关于位置参数模型(1.3)于 1964 年在文献[33]中引进的. 1973 年,Huber 在文献[34]中又将这种估计拓展到一般的线性模型(1.1). 自那以后,这一方向吸引了一些统计学者的注意,发表了不少论文.

M 估计的命名,与它同极大似然估计(Maximum Likelihood Estimate,缩写为 MLE)的联系有关. 设在模型(1.1)中,误差 $e_1,\cdots,e_n$ 独立同分布(以后简记为 iid.)有已知的公共密度 f,则似然函数为

$$L(\beta) = \prod_{i=1}^{n} f(Y_i - x_i'\beta).$$

β 的 MLE β_n^* 使 $L(\beta)$达到最大. 如果记 $\rho = -\log f$,则 β_n^* 就是 $\sum_{i=1}^{n}\rho(Y_i - x_i'\beta)$的最小值点. 所以 MLE 是 M 估计的一个特例,这个"M"也就是取自 MLE 的首位字母. 例如:

1) 若 $f(x)$为正态 $N(0,\sigma^2)$的密度,则

$$\rho(u) = (2\sigma^2)^{-1}u^2 + \text{const.},$$

引出的估计就是最小二乘估计(Least Squares Estimate,缩写为 LSE).

2) 若 $f(x)$为重指数分布(也称 Laplace 分布)的密度:

$$f(x) = (2\sigma)^{-1}\exp(-|x|/\sigma),$$

则

$$\rho(u) = |u|/\sigma + \text{const.}.$$

引出的估计是最小一乘估计或称最小绝对偏差估计(Least Absolute Deviation Estimate,缩写为 LADE).

LSE 和 LADE 是 M 估计的两个最重要的特例. 前者历史悠久,在很长时期,很大程度上可以说直到如今,在线性回归分析中占据着中心位置. LADE 也有悠久历史,在相当长的时期内,由于在计算及理论上的困难,进展不大. 近几十年来,随着计算问题的解决,在应用上愈来愈受人重视,在理论上(主要是大样本性质的)也有了比较深入的进展.

在 M 估计定义中的函数 ρ 在一定限度内可以自由选择,以适应不同的需要. 例如,当$|u|\to\infty$,$\rho(u)$增长较慢时,β_n 有较好的样本稳健性,即对样本中可能混入的少量异常值有较强的抵抗力. 例如 LADE 的稳健性优于 LSE 的稳健性,这种对稳健性的追求是引进 M 估计的动力. 不过要注意:稳健性与其说是一个优点,不如说是一个性质. 并不是愈稳健愈好,过于强调稳健性可能导致效率的损失. 另外,如果我们对误差分布类有相当的知识,则它也可用于指导我们选择适当的 ρ,以期达到更高的效率. 不过 ρ 的选择并

非完全自由，有其分析上和统计上的限制. 在以下几段我们就致力于讨论这个问题.

三、M 估计的存在定理

前面用(1.5)式定义了模型(1.1)中的 β_0的 M 估计 $\hat{\beta}_n$，故首先需要回答的问题，就是在何种条件下(1.5)右端的极值能够达到. 关于这个问题，有下面的一般结果：

定理1.1　设以下三条件成立：

1°　ρ 在 R^1处处连续.

2°　存在 a，使 $\rho(u)$当 $u\geqslant a$ 时，非降且不恒等于一常数；当 $u\leqslant a$ 时，非增且不恒等于一常数.

3°　矩阵 $X=(x_1\cdots x_n)$的秩为 p.

则存在 $\hat{\beta}_n\in R^p$，使(1.5)式成立.

条件3°是必要的，因若 X 的秩小于 p，则在模型(1.1)之下 β_0为不可估，估计 β_0的问题就失去了意义. 前两个条件只涉及 ρ，其意义可以这样理解：如果用 β 估计 β_0，则 $Y_i-x_i'\beta$ 可视为实际数据与这一估计之间的偏离，而 $\rho(Y_i-x_i'\beta)$可视为由这种偏离而引起的某种损失的度量. 在这一看法下，条件1°无非是说：损失随偏离连续变化. 对条件2°，设 $a=0$，则该条件意味着：同一个方向的偏离，绝对值愈大，所导致的损失也应愈大，至少不能更小. 从这个观点看，这两条要求都是自然的、起码的要求(在实用中所用的 ρ，其条件2°中的 a 都是0. 作为一个分析定理，我们把这个条件提得更一般一点).

现在进行定理的证明：记

$$H_0=\inf\{H_\rho(\beta):\beta\in R^p\},$$

按条件2°，对一切 $\beta\in R^p$，有 $H_\rho(\beta)\geqslant n\rho(a)$，故 $-\infty<H_0<\infty$. 在 R^p 中找一列点$\{b_m\}$，使

$$\lim H_\rho(b_m)=H_0.$$

若

$$\lim\|b_m\|=\infty \tag{1.6}$$

不成立，则可找到$\{b_m\}$的子列$\{b_{m'}\}$，使 $b_{m'}\to b\in R^p$. 利用 ρ 的连续

性，有 $H_\rho(b)=H_0$，于是 b 达到了 $H_\rho(\beta)$ 的最小值. 因此，为证定理1.1，只须证明(1.6)不可能. 用反证法：设(1.6)成立，则可抽出 $\{b_m\}$ 的一子列——为方便计，就设此子列即 $\{b_m\}$ 本身，使当 $m\to\infty$ 时，以下的断言成立：

i) $b_m/\|b_m\|\to$ 某单位向量 h.

ii) $x_i'b_m$ 有极限 g_i $(1\leqslant i\leqslant n)$，且

$$g_i=-\infty \quad (\text{当}1\leqslant i\leqslant n_1);$$

$$g_i=\infty \quad (\text{当}\ n_1+1\leqslant i\leqslant n_2);$$

$$|g_i|<\infty \quad (\text{当}\ n_2+1\leqslant i\leqslant n).$$

(为满足 ii)，必要时可更改 $x_1,\cdots,x_n$ 足标的编号.)必有：

$$x_i'h = 0 \quad (n_2 + 1 \leqslant i \leqslant n). \tag{1.7}$$

因若对此范围内的某个 i(1.7)不成立，则对这个 i 有：

$$|x_i'b_m| = \|b_m\|\,|x_i'b_m/\|b_m\|| \to \infty \cdot |x_i'h| = \infty.$$

这与 ii)矛盾.

根据(1.7)及 X 的秩为 p 的假定，得知 $x_i'h$ $(1\leqslant i\leqslant n_2)$不能全为0. 记

$$q = \min(\rho(\infty),\rho(-\infty)) - \rho(a).$$

按假定2°，有 $q>0$，因此根据 b，可找到充分大的 m，使

$$\sum_{i=n_2+1}^{n}\rho(Y_i - x_i'b_m) < \sum_{i=n_2+1}^{n}\rho(Y_i - g_i) + q/2; \tag{1.8}$$

$$Y_i - x_i'b_m > a \quad (\text{当}\ 1\leqslant i \leqslant n_1);$$

$$Y_i - x_i'b_m < a \quad (\text{当}\ n_1 + 1 \leqslant i \leqslant n_2).$$

记

$$d_i = |x_i'h|/|Y_i - x_i'b_m - a| \quad (1 \leqslant i \leqslant n_2).$$

因 $x_i'h$ $(1\leqslant i\leqslant n_2)$不全为0，故 $d\equiv\max(d_1,\cdots,d_{n_2})>0$. 为方便计且不失普遍性，设 $d=d_1$，且 $x_1'h>0$. 令 $\tilde{b}_m=b_m+h/d$，则

$$x_i'\tilde{b}_m= x_i'b_m \quad (n_2 + 1 \leqslant i \leqslant n),$$

$$Y_1 - x_1'\tilde{b}_m= Y_1 - x_1'b_m - x_1'h/d = a,$$

$$Y_i - x_i'\tilde{b}_m= Y_i - x_i'b_m - x_i'h/d$$

$$\geqslant Y_i - x_i'b_m - |x_i'h|/d_i$$
$$\geqslant Y_i - x_i'b_m - |Y_i - x_i'b_m - a|$$
$$= Y_i - x_i'b_m - (Y_i - x_i'b_m - a)$$
$$= a \quad (2 \leqslant i \leqslant n_1).$$

类似可证明

$$Y_i - x_i'\bar{b}_m \leqslant a \quad (n_1 + 1 \leqslant i \leqslant n_2).$$

综合以上诸式，并利用条件2°和(1.8)式，得

$$H_\rho(\bar{b}_m) \leqslant \rho(a) + (n_1 - 1)\rho(\infty) + (n_2 - n_1)\rho(-\infty) + \sum_{i=n_2+1}^{n} \rho(Y_i - x_i'b_m)$$
$$\leqslant n_1\rho(\infty) + (n_2 - n_1)\rho(-\infty) + \sum_{i=n_2+1}^{n} \rho(Y_i - g_i) - q/2.$$

但由 ρ 的连续性和$\{b_m\}$所满足的条件 ii)，以及 $H_\rho(b_m) \to H_0$的事实知：上式右端前三项之和即为 H_0. 于是由上式得到矛盾的结果 $H_\rho(\bar{b}_m) < H_0$. 这证明了(1.6)不可能. 即证明了本定理. ▎

四、从统计的观点看 ρ 应满足的条件

上段所述的是：为了 $\hat{\beta}_n$ 存在，ρ 要满足的条件. 但 $\hat{\beta}_n$ 存在，并不等于它在统计上合理. 为了得到"在统计上合理"的 M 估计，ρ 应满足什么条件呢？本段的目的即是提出和解释这样的条件. 注意：此处说的是"解释"而非"证明". 因为什么是"统计上合理"的估计，本来就无严格的定义，只能通过一些启发性的解释来说明这样的条件是应当满足的.

考虑模型(1.1)，定义

$$h_i(t) = E\rho(e_i + t).$$

这里暂且假定 $E|\rho(e_i+t)| < \infty$，对任何 $t \in R^1$，以使 $h_i(t)$有意义. 所要提出的条件是：

$$\text{当 } t \neq 0 \text{ 时}, h_i(t) > h_i(0) \quad (i = 1, 2, \cdots), \tag{1.9}$$

换句话说，$h_i(t)$以 $t=0$为唯一的最小值点.

下面对条件(1.9)作一个启发性的解释：不论用什么方法估计 β_0，总是希望估计值能有更多的机会接近 β_0，今若设想就以 β_0 估计 β_0(这是最理想的)，平均损失为

$$EH_\rho(\beta_0) = \sum_{i=1}^{n} E\rho(e_i) = \sum_{i=1}^{n} h_i(0),$$

而若以某个非 β_0 的 β 去估计 β_0，则平均损失为

$$EH_\rho(\beta) = \sum_{i=1}^{n} h_i[x_i'(\beta_0 - \beta)].$$

由于 $\beta_0-\beta\neq 0$，若 $(x_1\cdots x_n)$ 的秩为 p，则 $x_i'(\beta_0-\beta)$ $(1\leqslant i\leqslant n)$ 不全为0. 按(1.9)，应有：

$$EH_\rho(\beta) > EH_\rho(\beta_0). \tag{1.10}$$

一般而言，(1.10)式意味着有更多的可能是

$$H_\rho(\beta) > H_\rho(\beta_0).$$

因为 M 估计使 $H_\rho(\beta)$ 达到最小，上式意味着使 $H_\rho(\beta)$ 达到最小的点，即 M 估计 $\hat{\beta}_n$ 应有更多的机会在真值 β_0 附近，而不是在远离真值 β_0 的某个地方. 反之，若(1.9)不满足，则相反的情况就可能发生. 这解释了条件(1.9)的必需性质.

当然，这一解释过于笼统，不一定能使所有的人满意. 在第2章及以后章节中我们将看到在有关 M 估计的许多大样本定理的假定中，都隐含这个条件. 以下我们针对最简单的模型(1.3)及 LSE 和 LADE 这两个实例，可以更清楚地看到这个条件的意义. 为简单计，设 e_i 为 iid..

先考虑 α_0 的 LSE，它就是 $Y_1,\cdots,Y_n$ 的样本均值 $\overline{Y}_n$，按(1.3)，有

$$\overline{Y}_n = \alpha_0 + \bar{e}_n, \bar{e}_n = \sum_{i=1}^{n} e_i/n.$$

对 LSE 而言，条件(1.9)等价于 $Ee_i=0$. 若这个条件满足，则按大数律有 $\bar{e}_n \xrightarrow{P} 0$，因而 $\overline{Y}_n \xrightarrow{P} \alpha_0$，即 $\overline{Y}_n$ 为 α_0 的相合估计. 反之，若 $h(t)=E(e_1+t)^2$ 在 $t=a$ 处达到最小值，而 $a\neq 0$，则 $Ee_1=-a$，而 $\overline{Y}_n \xrightarrow{P} \alpha_0-a\neq\alpha_0$，即 $\overline{Y}_n$ 不为相合. 因此，在本例中 M 估计 $\overline{Y}_n$ 是否

相合取决于条件(1.9)是否满足. 因为相合性是对一个估计的最起码的要求,所以条件(1.9)不可少.

其次看 α_0的 LADE,它就是 $Y_1,\cdots,Y_n$ 的样本中位数 $\hat{m}_n$. 对 LADE 而言,条件(1.9)等价于 e_1有唯一的中位数0. 若这个条件满足,则易证明(参见文献[3])

$$\hat{m}_n \xrightarrow{P} \operatorname{med}(Y_1) = \alpha_0,$$

即 $\hat{m}_n$ 为相合的. 同样,若 $h(t)=E|e_1+t|$ 的唯一最小值在 $a\neq 0$处达到,或者 $h(t)$有不止一个的最小值点,则 $\hat{m}_n$ 不为相合. 所以在本例中,条件(1.9)也不能少.

读者或许会问:由于一般地我们并不确切知道 e_i 的分布,因而也就求不出 $h_i(t)$. 这样,我们如何能知道条件(1.9)是否满足呢?以下我们从几个方面来回答这个问题.

1) 在实际应用中,回归函数按应用的目的,总是有其特定含义的. 例如在(1.1)中 $x_i'\beta_0$可能是 Y_i 的期望(均值回归),这个含义决定了 $Ee_i=0$,因此也就决定了若取 $\rho(u)=u^2$,则条件(1.9)必满足. 同样,若 $x_i'\beta_0$是 Y_i 的中位数(中位回归),则这个含义决定了 $\operatorname{med}(e_i)=0$,若假定 e_i 的中位数唯一,则在取 $\rho(u)=|u|$时,条件(1.9)满足. 一般来说,若 $x_i'\beta_0$为 Y_i 的 p 分位数($0<p<1$),则 e_i 的 p 分位数为0. 若假定 e_i 的 p 分位数唯一,则易证明,当取

$$\begin{aligned}\rho(u) &= p|u| \quad (u \geqslant 0)\\ &= (1-p)|u| \quad (u<0)\end{aligned} \tag{1.11}$$

时,条件(1.9)满足. 换句话说,回归函数的含义指导我们选择 ρ,以使(1.9)成立. 在此要提醒的是:不要将此看法极端化,以为回归函数的含义总是完全决定 ρ 的选择,因而就失掉灵活性. 举一个例子:在许多情况下,有理由假定 Y_i 服从对称分布,而 $x_i'\beta_0$为对称中心. 这时,$x_i'\beta_0$既是 Y_i 的期望,也是其中位数,因此 $\rho(u)=u^2$和 $\rho(u)=|u|$都能满足(1.9). 还有其他可能的选择,见以下第3条.

2) 对一般含常数项的模型(1.2),有时在实用上感兴趣的只是回归系数 β_0的估计,α_0却不重要. 这时,条件(1.9)可放宽为:

$h_i(t)$ 有唯一的最小值点 a，a 不必为0. 因为若令 $\tilde{\alpha}_0=\alpha_0-a$，$\tilde{e}_i=e_i+a$，而将(1.2)改写为

$$Y_i=\tilde{\alpha}_0+x_i'\beta_0+\tilde{e}_i,$$

则
$$\tilde{h}_i(t)\equiv E\rho(\tilde{e}_i+t)=E\rho(e_i+a+t).$$

后者在 $t=0$ 处有唯一最小值点 $t=0$，因此条件(1.9)满足. 在新的模型中，β_0 的值依旧，其估计不受影响，但如不知道 a，则 $\tilde{\alpha}_0$ 无法还原为 α_0，因而后者的估计会出现系统偏差.

3) 在实用上，ρ 多选择为偶函数. 这时，若 e_i 的分布关于0对称，且有密度，则在一般的条件下，可以证明条件(1.9)成立.

定理1.2 设 ρ 为偶函数，在 $[0,\infty)$ 非降；随机变量 e 有密度 f，且 f 为偶函数，在 $[0,\infty)$ 非增. 若以下三个条件满足其一，则 $E\rho(e+t)$ 以 $t=0$ 为唯一最小值点.

1° ρ 在 $[0,\infty)$ 严增；

2° f 在 $[0,\infty)$ 严降，ρ 不恒等于常数；

3° 存在 $\Delta>0$，使 ρ 在 $[0,\Delta)$ 严增；f 在 $[0,\Delta)$ 严降.

证 记

$$J(t)=\int_{-\infty}^{\infty}(\rho(u+t)-\rho(u))f(u)\mathrm{d}u, \tag{1.12}$$

因为 ρ、f 皆为偶函数，易知 J 为偶函数，故只须考察 $t>0$ 的情况. 把(1.12)式右端的积分拆成 $\int_{-\infty}^{0}+\int_{0}^{\infty}$，在后一积分中作变数代换 $v=-u-t$，并仍改 v 为 u. 利用 ρ、f 的偶性，易得

$$\begin{aligned}J(t)=&\int_0^{\infty}(\rho(u+t)-\rho(u))(f(u)-f(u+t))\mathrm{d}u\\&+\int_{-\infty}^{0}(\rho(u)-\rho(u+t))f(u+t)\mathrm{d}u\\\equiv&J_1(t)+J_2(t).\end{aligned} \tag{1.13}$$

易见 $J_1(t)\geqslant0$，且若1°、2°满足其一，则 $J_1(t)>0$. 对 $J_2(t)$ 作变量代换 $v=u+t/2$，得

$$J_2(t)\geqslant\int_{-t/2}^{t/2}(\rho(v-t/2)-\rho(v+t/2))f(v+t/2)\mathrm{d}v$$

$$= \int_0^{t/2} + \int_{-t/2}^0 = J_3(t) + J_4(t).$$

在 $J_4(t)$ 中作代换 $v = -u$，并仍改积分变量 u 为 v，利用 ρ 和 f 的偶性，把变换后的 $J_4(t)$ 与 $J_3(t)$ 结合起来，得到

$$J_2(t) \geqslant \int_0^{t/2} (\rho(v + t/2) - \rho(v - t/2))(f(v - t/2) - f(v + t/2))\mathrm{d}v.$$

从对 ρ 和 f 的全部假定，易知 $J_2(t) \geqslant 0$，且若条件3°满足，则 $J_2(t) > 0$. 将此与前面已证的 $J_1(t)$ 的性质结合起来，即知：对于任何 $t > 0$，有 $J(t) > 0$. 即证明了所要的结果. ▌

前已提到，条件(1.9)隐含了要求 $E|\rho(e_i + t)| < \infty$. 有时，把这条件改写为以下形式：

$$E(\rho(e_i + t) - \rho(e_i)) > 0 \quad (\text{当 } t \neq 0), \tag{1.14}$$

它只要求

$$E|\rho(e_i + t) - \rho(e_i)| < \infty \quad (\text{对任何 } t \in R^1). \tag{1.15}$$

通常，(1.15)比 $E|\rho(e_i + t)| < \infty$ 要弱. 如对 LSE，按(1.9)的形状要求 $Ee_i^2 < \infty$；而按(1.15)，则只须 $E|e_i| < \infty$. 对 LADE，按(1.9)的形状，要求 $E|e_i| < \infty$；而按(1.15)，则无任何要求.

五、ρ 为凸函数的情况

这个情况在本书中起着重要的作用. 这里先介绍一些本书需要的初步知识. 在后面第4、5章中用到的进一步知识，将在第4章的附录中介绍.

定义于 R^p 的函数 f 称为**凸函数**，如果对任何 $u_1 \in R^p$，$u_2 \in R^p$ 及 $0 < \lambda < 1$，必有

$$f(\lambda u_1 + (1 - \lambda)u_2) \leqslant \lambda f(u_1) + (1 - \lambda)f(u_2).$$

如果当 $u_1 \neq u_2$ 及 $0 < \lambda < 1$ 时，上式总有严格不等号成立，则称 f 为**严凸的**.

定理1.3 设 ρ 为定义于 R^1 的凸函数，则以下的一些结论成立：

1° 若 ρ 是非单调的，则它必满足定理1.1的条件1°和2°，且

$\liminf_{|u|\to\infty} \rho(u)/|u|>0$.

2° 按(1.4)定义 $H_\rho(\beta)$，则 $H_\rho(\beta)$ 为 β 的凸函数. 如果 ρ 为严凸，且矩阵 $X=(x_1\cdots x_n)$ 的秩为 p，则 $H_\rho(\beta)$ 也是严凸的. 又若 ρ 非单调，且 X 的秩为 p，则 $\lim_{\|\beta\|\to\infty} H_\rho(\beta)=\infty$.

3° 若 S 是 R^p 中的一个封闭曲面，将 R^p 分成内外两部分，分别记为 A、B. 若 β_0 是 A 的一个内点，并且 $\inf\{H_\rho(\beta):\beta\in S\}>H_\rho(\beta_0)$，则当 $\beta\in B$ 时，必有 $H_\rho(\beta)>H_\rho(\beta_0)$.

证 1° 由 ρ 的凸性可推出其连续性. 由于 ρ 为非单调，存在 $a<b$，使 $0<\rho(a)<\rho(b)$. 这时，根据凸性易知，有

$$\rho(u)\geqslant\rho(b)+(b-a)^{-1}(\rho(b)-\rho(a))(u-b)\quad(u\geqslant b).$$

由此知

$$\liminf_{u\to\infty}\rho(u)/|u|\geqslant(b-a)^{-1}(\rho(b)-\rho(a))>0.$$

类似地，由 ρ 为非单调，存在 $c<d$，使 $\rho(c)>\rho(d)>0$. 由此出发可证 $\liminf_{u\to-\infty}\rho(u)/|u|>0$. 即证明了

$$\liminf_{|u|\to\infty}\rho(u)/|u|>0. \tag{1.16}$$

由(1.16)知 $\rho(\pm\infty)=\infty$，据此及 ρ 的连续性知，ρ 必在某点 a_0 达到最小值. 再由凸性即知，当 $u\geqslant a_0$ 时，ρ 非降；当 $u\leqslant a_0$ 时，ρ 非增. 即证明了1°.

2° $H_\rho(\beta)$ 的凸性是显然的. 若 ρ 为严凸，且 $X\equiv(x_1\cdots x_n)$ 有秩 p，则对 R^p 中任何 $\beta_1\neq\beta_2$，必存在 x_j $(1\leqslant j\leqslant n)$，使 $x_j'\beta_1\neq x_j'\beta_2$. 于是对任何 $\lambda\in(0,1)$，有

$H_\rho(\lambda\beta_1+(1-\lambda)\beta_2)=\sum_{i=1}^n\rho(\lambda(Y_i-x_i'\beta_1)+(1-\lambda)(Y_i-x_i'\beta_2))$，对任何 $i=1,\cdots,n$，有

$$\rho(\lambda(Y_i-x_i'\beta_1)+(1-\lambda)(Y_i-x_i'\beta_2))$$
$$\leqslant\lambda\rho(Y_i-x_i'\beta_1)+(1-\lambda)\rho(Y_i-x_i'\beta_2),$$

且因 $Y_j-x_j'\beta_1\neq Y_j-x_j'\beta_2$，当 $i=j$ 时，上式有严格不等号成立，故

$$H_\rho(\lambda\beta_1+(1-\lambda)\beta_2)<\lambda H_\rho(\beta_1)+(1-\lambda)H_\rho(\beta_2),$$

从而证明了 $H_\rho(\beta)$ 的严凸性.

当 ρ 非单调时，上面已证明 $\rho(\pm\infty)=\infty$. 若 X 的秩为 p，则 $XX'>0$，因而当 $\|\beta\|\to\infty$ 时，有 $\beta'XX'\beta\to\infty$，故 $\max\limits_{1\leqslant i\leqslant n}\|x_i'\beta\|\to\infty$，因为 ρ 具有有限的下界，知

$$\lim_{\|\beta\|\to\infty} H_\rho(\beta) = \infty.$$

即证明了2°.

3°的证明很简单：如果存在 $\beta\in B$ 使 $H_\rho(\beta)\leqslant H_\rho(\beta_0)$，则由 $H_\rho(\beta)$ 的凸性知，在 β_0 和 β 连成的线段 l 上的每点 β_1 都有 $H_\rho(\beta_1)\leqslant H_\rho(\beta_0)$. 但 l 与 S 有交点 β_2，故 $H_\rho(\beta_2)\leqslant H_\rho(\beta_0)$. 这与 $\inf\{H_\rho(\beta):\beta\in S\}>H_\rho(\beta_0)$ 矛盾. 定理证毕. ▍

本定理的1°和2°说明了：当 ρ 为凸且非单调时，由(1.5)定义的 M 估计 $\hat{\beta}_n$ 必存在(当然，需要矩阵 X 的秩为 p 这个条件)，且若 ρ 为严凸，则 $\hat{\beta}_n$ 唯一. (1.16)说明了：当 $|u|\to\infty$ 时，$\rho(u)$ 的增长速度至少与 $|u|$ 一样. 这表明，当要求稳健性很强时，使用凸的 ρ 也许是不合适的. 本定理的3°虽然很简单，但在以后章节中起着重要的作用，甚至可以说，ρ 为凸函数这种情况之所以在 M 估计理论中占如此重要的地位，在相当程度上，是与这一简单性质有关的.

下面考察 ρ 为凸这一性质与条件(1.9)的关系. 先引入几个记号，因为 ρ 为凸，在每个 $u\in R^1$ 处，ρ 的左、右导数皆存在，分别记之为 $\Psi_-(u)$ 和 $\Psi_+(u)$. 因为 ρ 为凸，Ψ_+、Ψ_- 都非降，确切地说，当 $a<b$ 时，

$$\Psi_-(a)\leqslant\Psi_+(a)\leqslant\Psi_-(b)\leqslant\Psi_+(b).$$

因此，若以 Ψ 记任一满足条件

$$\Psi_-(u)\leqslant\Psi(u)\leqslant\Psi_+(u)\quad(u\in R^1)\tag{1.17}$$

的函数，则 Ψ 也必为非降的.

定理1.4 设 ρ 为 R^1 上的凸函数，Ψ_+、Ψ_- 和 Ψ 的意义同上，又 e 为随机变量. 则

$$E|\rho(e+t)-\rho(e)|<\infty\quad(t\in R^1),\tag{1.18}$$

且

$$g(t)\equiv E(\rho(e+t)-\rho(e))\quad(t\in R^1)\tag{1.19}$$

以0为唯一最小值点的充要条件是:对某个满足(1.17)的 Ψ,有

$$E|\Psi(e+t)|<\infty \quad (t\in R^1), \tag{1.20}$$

$$E\Psi(e)=0, \tag{1.21}$$

$$tE\Psi(e+t)>0 \quad (\text{当 } t\neq 0). \tag{1.22}$$

证 先证(1.18)与(1.20)等价. 以 F 记 e 的分布,设(1.18)成立,固定 t,根据 ρ 的凸性,易见

$$\rho(u+t+1)-\rho(u+t)\geqslant\Psi(u+t)$$
$$\geqslant\rho(u+t)-\rho(u+t-1),$$

因此

$$|\Psi(e+t)|\leqslant|\rho(e+t+1)-\rho(e+t)|$$
$$+|\rho(e+t)-\rho(e+t-1)|.$$

由这个不等式及(1.18),即得(1.20). 反过来,设(1.20)对某个满足(1.17)的 Ψ 成立,易见(1.20)对任何满足(1.17)的 $\tilde{\Psi}$ 都成立. 事实上,因为 Ψ、$\tilde{\Psi}$ 都满足(1.17),有

$$\Psi(u+t-1)\leqslant\tilde{\Psi}(u+t)\leqslant\Psi(u+t+1),$$

故 $|\tilde{\Psi}(e+t)|\leqslant|\Psi(e+t+1)|+|\Psi(e+t-1)|$. 因为 Ψ 满足(1.20),知 $\tilde{\Psi}$ 也如此. 根据 ρ 的凸性,有

$$\rho(u+t)-\rho(u)>0\Rightarrow\rho(u+t)-\rho(u)$$
$$\leqslant|t|\max(|\Psi_-(u-t)|,|\Psi_+(u+t)|),$$
$$\rho(u+t)-\rho(u)<0\Rightarrow|\rho(u+t)-\rho(u)|$$
$$\leqslant|t|\max(|\Psi_-(u)|,|\Psi_+(u)|). \tag{1.23}$$

由此,利用(1.20)式对任何满足(1.17)的 Ψ 皆成立的事实,即得(1.18).

现设(1.20)~(1.22)成立,往下证明0为 $g(t)$ 的唯一最小值点. 由(1.17),有

$$E\Psi_+(e)\geqslant 0\geqslant E\Psi_-(e). \tag{1.24}$$

因为 ρ 为凸,易知(1.19)式定义的 g 也为凸. 现设 $0<t<1$,则在(1.23)第一式中,$\Psi_+(u+t)\geqslant 0$,故

$$0\leqslant\Psi_+(u+t)\leqslant\Psi_+(u+1),$$

又

$$|\Psi_-(u+t)| \leqslant \max(|\Psi_-(u)|,\ \Psi_-(u+1)|),$$

据此，由(1.23)式，得

$$|\rho(u+t)-\rho(u)|/t \leqslant |\Psi_+(u+1)| + |\Psi_-(u+1)| + 2|\Psi_-(u)| + |\Psi_+(u)|.$$

因为上面已证明(1.20)对任何满足(1.17)的 Ψ 都成立，根据上式，用控制收敛定理，即得

$$\begin{aligned} g'_+(0) &\equiv \lim_{t\downarrow 0} E[t^{-1}(\rho(e+t)-\rho(e))] \\ &= E[\lim_{t\downarrow 0} t^{-1}(\rho(e+t)-\rho(e))] \\ &= E\Psi_+(e) \geqslant 0. \end{aligned} \tag{1.25}$$

类似地(取$-1<t<0$，并令 $t\uparrow 0$)可证明

$$g'_-(0) = E\Psi_-(e) \leqslant 0.$$

据此，由 g 的凸性，即知0为 g 的最小值点. 以下证唯一性：设某个 $t_0>0$也是 g 的最小值点. 根据 g 的凸性，$(0,t_0)$内任一点都是 g 的最小值点，故 $g'(t)=0\ (0<t<t_0)$. 按导出(1.25)的道理，可得

$$g'_+(t) = E\Psi_+(e+t),\ g'_-(t) = E\Psi_-(u+t),$$

故

$$E\Psi_+(e+t) = E\Psi_-(e+t) = 0 \quad (0<t<t_0). \tag{1.26}$$

由(1.17)及(1.26)，知

$$E\Psi(e+t) = 0 \quad (0<t<t_0).$$

这与(1.22)矛盾. 当 $t_0<0$时，证法类似.

反过来，设(1.18)成立，并且0是 g 的唯一最小值点，则有 $g'_+(0)\geqslant 0\geqslant g'_-(0)$. 由于前面已证明(1.20)，故(1.25)(及类似关于 $g'_-(0)$)式成立，因而

$$E\Psi_+(e) \geqslant 0 \geqslant E\Psi_-(e). \tag{1.27}$$

根据(1.27)，即知存在 Ψ 适合(1.17)，使(1.21)成立. 为证(1.22)，用反证法：设存在 $t_0>0$，使

$$E\Psi(e+t_0) \leqslant 0. \tag{1.28}$$

因为 Ψ 非降，由(1.21)及(1.28)，知 $E\Psi(e+t_0)=0$. 据此，由 Ψ 的单调性，易知

$$E\Psi_+(e+t)=E\Psi_-(e-t)=0\quad(0<t<t_0).$$

利用此式，根据推导(1.25)的论据，即得 $g'(t)=0$（当$0<t<t_0$）. 因为 g 为凸，知$(0,t_0)$内每点都是 g 的最小值点. 这与 g 有唯一最小值点矛盾. $t_0<0$的情况，证法类似. 定理证毕. ▌

这个定理把条件(1.9)转化到了对函数 Ψ 的要求. 在以后各章，我们会遇到一些定理，其条件通过 Ψ 来表达. 初始一看，这种条件的统计意义不明显. 有了定理1.4，就可以看出，那些条件保证了(1.9)这个必需的条件. 不过，为了特定的目的，所提的条件往往超出了这个范围，今后我们会有机会讨论这一点.

在关于条件(1.9)的讨论中，我们曾指出：若 g 有唯一最小值点 a，但 $a\neq 0$. 则除了模型中的常数项外，对回归系数的估计并无影响. 下一定理表明：若 ρ 为严凸且非单调，则这一点总是成立的.

定理1.5 若 ρ 为非单调的严凸函数，且(1.18)成立. 则由(1.19)式定义的 g，必有唯一最小值点.

证 由 ρ 为严凸知 g 也为严凸，故如有最小值点，则必唯一. 往下证 $g(\pm\infty)=\infty$，由此就推出 g 必有最小值点，因而即可证明本定理.

因为 ρ 为凸且非单调，按定理1.3的1°，有 $\rho(\pm\infty)=\infty$. 根据定理1.4，由(1.18)知(1.20)成立. 又根据 $\rho(\pm\infty)=\infty$及 ρ 的凸性知，存在 u_0及 $\varepsilon\in(0,1)$，使

$$\Psi_+(u)\geqslant\varepsilon\quad(\text{当 } u\geqslant u_0).$$

以 F 记 e 的分布. 找 a，使

$$\int_{-\infty}^{a}|\Psi_+(u)|\mathrm{d}F(u)<\varepsilon^2/3,\ \int_{-\infty}^{a}\mathrm{d}F(u)<1-\varepsilon/2.$$

记 $t_0=\max(0,u_0-a)$，因 $t_0\geqslant 0$，当 $t\geqslant t_0$时，有

$$\int_{-\infty}^{a}\Psi_+(u+t)\mathrm{d}F(u)\geqslant\int_{-\infty}^{a}\Psi_+(u)\mathrm{d}F(u)\geqslant-\varepsilon^2/3.$$

当 $u\geqslant a$，$t\geqslant t_0$时，有 $u+t\geqslant u_0$. 故

$$\int_{a}^{\infty}\Psi_+(u+t)\mathrm{d}F(u)\geqslant\varepsilon\int_{a}^{\infty}\mathrm{d}F(u)\geqslant\varepsilon^2/2.$$

将二式相加，得

$$\int_{-\infty}^{\infty} \Psi_+ (u+t) \mathrm{d}F(u) = E\Psi_+ (e+t) \geqslant \varepsilon^2/6 \quad (t \geqslant t_0). \tag{1.29}$$

因为 Ψ 非降，因而 $E\Psi_+(e+t)$ 也非降. 由(1.29)可知，当 $t\to\infty$ 时，$E\Psi_+(e+t)$ 的极限存在，且大于0(可为∞)，以 l 记这一极限(如为∞，则取 $l=1$)，按导出(1.25)的方法，得

$$g'_+ (t) = E\Psi_- (e+t) \geqslant \varepsilon^2/6 > 0 \quad (t \geqslant t_0).$$

根据此，由 g 的凸性，即得 $g(\infty)=\infty$. 类似地可证明 $g(-\infty)=\infty$. 定理证毕. ▌

注意：此处我们并未假定对任何 $t\in R^1$，有 $E|\rho(e+t)|<\infty$. 如作了这一假定，则由 Fatou 引理即可推出 $g(\pm\infty)=\infty$.

§1.2 M 估计作为方程的解

设有线性模型(1.1)，用(1.5)定义 β_0 的 M 估计 $\hat{\beta}_n$，并设 ρ 的导数 $\rho'=\Psi$ 在 R^1 上处处存在. 则 $\hat{\beta}_n$ 必满足方程

$$\sum_{i=1}^{n} \Psi(Y_i - x_i'\beta) x_i = 0. \tag{1.30}$$

基于这一点，有的工作(参见文献[51]、[52]、[67]等)就脱开(1.30)与(1.5)的关系，而径直从(1.30)出发，即给定一个定义于 R^1 的函数 Ψ，然后把 β_0 的 M 估计定义为方程(1.30)的解. 这也可以看成是 M 估计的另一种定义(独立于(1.5)的定义). 因为除非 ρ' 处处存在，由(1.5)出发的定义不能是(1.30)的解. 反之，除非 Ψ 满足一系列的条件，(1.30)的解也不可能对某个 ρ 适合(1.5).

从基于定义(1.30)出发的工作分析，可看出其动机在于：若假定 Ψ 有适当的光滑性，则按(1.30)左端作 Taylor 展开，提供了对 $\hat{\beta}_n$ 作大样本分析的一个途径. 在较早的时候，没有找到除此以外的有效方法，而直接从(1.5)出发又不对 ρ 作充分光滑的假定，在处理上很困难. 以后的工作表明：对一些基本的大样本问题——如 $\hat{\beta}_n$ 的相合性与渐近正态性，不须对 ρ 作多少光滑性假定，就可以

解决得很好.在这些情况下,从(1.30)出发,等于在较强的假定下,推出同样或更弱的结果.例如,从(1.30)出发,连LADE这样重要的情况也无法对付.

这样,我们本来可以结束关于这个定义的讨论(事实上,本书从不依(1.30)的方式定义 M 估计),但由于从这一定义出发的讨论在文献中占有一定的数量,故我们应花一点篇幅来考察一下这个定义的主要问题——方程(1.30)的解的存在问题.基于定理1.1,不难证明以下的结果:

定理1.6 假定 Ψ 在 R^1 处处连续,且满足以下两条件之一:

1° 存在 a,使当 $u\geqslant a$ 时,$\Psi(u)\geqslant 0$,但不恒等于0;$u<a$ 时,$\Psi(u)\leqslant 0$,但不恒等于0.

2° Ψ 在 R^1 非降,且不是非负或非正.

则方程(1.30)必有至少一个解.

证 令 $\rho(u)=\int_a^u \Psi(v)\mathrm{d}v$. 若1°成立,则 ρ 在 R^1 连续,当 $u\geqslant a$ 时,非降;$u\leqslant a$ 时,非增.又因 Ψ 在 $u\geqslant a$ 及 $u<a$ 处都不恒等于常数,有

$$\rho(\infty)>0=\rho(a),\ \rho(-\infty)>0=\rho(a),$$

因此根据定理1.1知,$H_\rho(\beta)$(见(1.4))有最小值点 $\hat{\beta}_n$. 由 Ψ 连续,知 $\rho'=\Psi$ 处处成立,故 $\hat{\beta}_n$ 满足(1.30).

若2°成立,因 Ψ 处处连续,且不为非负或非正,故必存在 a,使 $\Psi(a)=0$. 再利用 Ψ 非降,且不是非负或非正,知 Ψ 满足1°,故由已证部分知,(1.30)的解存在.定理证毕. ▌

从以上证明可看出,2°其实是1°的特例.之所以单独列出,是因为对这个情况,由于 $\rho(u)=\int_a^u \Psi(v)\mathrm{d}v$ 所定义的 ρ 为凸的,因此进一步可推出这时(1.30)的解与(1.5)的解等价.且如 Ψ 为严增,则 ρ 为严凸.这时(1.5)和(1.30)都只有唯一的解.

迄今为止,还没有发现一种方法,能不依赖于通过令 $H_\rho(\beta)$ 对 β 的导数为0去得出方程(1.30),从而证明它有解.而只有做到了这一点,才能使 M 估计的方程的定义真正从极值定义中独立出

来.因此,在有些文献中从(1.30)出发的工作,其中对 Ψ 的假定包含了本定理中施加于 Ψ 的条件.也有的工作(例如文献[66])中对 Ψ 的假定与本定理的假定比较,互有出入,这时就不能保证(1.30)有解.特别是,Ψ 在 R^1 处处连续的条件很重要,那怕 Ψ 只在一个点不连续,也可以使(1.30)无解.让我们来看一个例子:

问题源出于文献[26],在其中作者研究(1.1)中 β_0的 M 估计,取函数 ρ 为

$$\rho(u) = \delta|u| + (1-\delta)u^2 \quad (0<\delta<1),$$

并用(1.5)定义 $\hat{\beta}_n$.这个 ρ 是 $|u|$(相当于 LADE)和 u^2(相当于 LSE)的混合,取法的用意大概是想综合二者的长处.文献[26]的作者想借助于方程(1.30),但麻烦的是:$\rho'(0)$不存在.因为 ρ' 只在一个点0不存在,而该点正是 ρ 的最小值点.看来规定 $\rho'(0)=0$是可行的,于是,令

$$\Psi(u) = \delta \cdot \mathrm{sgn}(u) + 2(1-\delta)u, \tag{1.31}$$

而把 $\hat{\beta}_n$ 定义为(1.30)的解.但作者忽略了在 Ψ 的这种规定下(1.30)是否有解的问题.这个问题在文献[16]中考察了,结果表明:即使对于最简单的情况——位置参数模型(1.3)以及 e_i 的公共分布满足很一般的条件下,(1.30)也可以无解.

定理1.7　设

$$Y_i = \alpha_0 + e_i \quad (1 \leqslant i \leqslant n, n \geqslant 1),$$

e_i 为 iid.,有公共密度 f,$f(c)>0$,且 f 在 c 点连续,其中 c 是函数

$$Q(\alpha) = \delta E|e_1 - \alpha| + (1-\delta)E|e_1 - \alpha|^2$$

的唯一最小值点(因为 ρ 为严凸,最小点必唯一).以 p_n 记事件{方程(1.30)无解}的概率,其中 Ψ 由(1.31)决定,则

$$\lim_{n\to\infty} p_n = \delta f(c)/(1-\delta+\delta f(c)). \tag{1.32}$$

当$0<\delta<1$时,此值大于0,因而 $\hat{\beta}_n$ 没有定义.注意:对此处考察的特殊模型,方程(1.30)有形式

$$\delta\sum_{i=1}^{n}\mathrm{sgn}(Y_i-\alpha) + 2(1-\delta)\sum_{i=1}^{n}(Y_i-\alpha) = 0. \tag{1.33}$$

这个定理的证明很长,有兴趣的读者可参看文献[16].在此处

我们考虑一个特殊情况：$e_1, e_2, \cdots$的公共分布为 $R(-1/2, 1/2)$——即区间$(-1/2,1/2)$上的均匀分布. 我们不证明(1.32)，而只证明一个较弱的形式：

$$\lim\inf p_n > 0. \tag{1.34}$$

由(1.34)同样得出 $\hat{\beta}_n$ 不可按方程(1.30)所定义的结论.

为证(1.34)，不妨设 $\alpha_0=1/2$. 这样，$Y_1, Y_2, \cdots$为 iid.，且有公共分布 $R(0,1)$. 为了书写简单起见，考虑 $\delta=2/3$的特例. 这时方程(1.33)简化为

$$\sum_{i=1}^{n}\operatorname{sgn}(Y_i-\alpha)+\sum_{i=1}^{n}(Y_i-\alpha)=0. \tag{1.35}$$

以 $\xi_1\leqslant\cdots\leqslant\xi_n$ 记 $Y_1,\cdots,Y_n$ 的次序统计量，(1.35)可写为

$$\sum_{i=1}^{n}\operatorname{sgn}(\xi_i-\alpha)+\sum_{i=1}^{n}(\xi_i-\alpha)=0. \tag{1.36}$$

记 $L_{n1}=[1/2-2/\sqrt{n}, 1/2]$，$L_{n2}=[1/2, 1/2+2/\sqrt{n}]$，

$$L_{n3}=[1/4-1/\sqrt{n}, 1/4+2/\sqrt{n}];$$

$$n_1=[n/2-\sqrt{n}],\ n_2=[n/2+\sqrt{n}],$$

其中$[a]$表示不超过 a 的最大整数. 定义

$$W_n=\sum_{i=1}^{n_2}\xi_i/n_2,\ V_n=\sum_{i=n_2+1}^{n}\xi_i \tag{1.37}$$

及事件 $A_i\equiv A_{in}=\{\xi_{n_i}\in L_{n_i}\}$ $(i=1,2)$，

$$B\equiv B_n=\{W_n\in[\xi_{n_2}/2-1/\sqrt{n}, \xi_{n_2}/2+1/\sqrt{n}]\}.$$

则易见

$$\{W_n\in L_{n3}\}\supset A_2\cap B. \tag{1.38}$$

往下证明

$$\lim\inf P(A_{1n}\cap A_{2n}\cap B_n)\equiv q>0. \tag{1.39}$$

事实上，由$[n/2\pm\sqrt{n}]/n=1/2\pm1/\sqrt{n}+o(1/\sqrt{n})$，按次序统计量的极限定理，有

$$\sqrt{n}\,(\xi_{[n/2\pm\sqrt{n}]}-(1/2\pm1/\sqrt{n}))\xrightarrow{\mathscr{L}}N(0,1/4).$$

由此立即得出

$$\lim P(A_{1n}) = \lim P(A_{2n}) = \Phi(2) - \Phi(-2). \tag{1.40}$$

此处及以下用 Φ 记 $N(0,1)$的分布函数.

考虑条件概率

$$P(B_n|u) \equiv P(B_n|\xi_{n_2} = u),$$

按次序统计量理论中周知的事实(例如文献[3]的定理1.3)知，在 $\xi_{n_2}=u$ 条件下，$\xi_1+\cdots+\xi_{n_2-1}$的条件分布就是 n_2-1个 iid. $R(0,u)$ 变量之和的分布. 因此，按 iid. 和逼近正态分布的 Berry-Esseen 理论(例如文献[49]第5章定理3)，存在与 n 和 $u\in(0,1)$都无关的常数 A，使对$-\infty<a<b<\infty$，有

$$\left|P\left(a \leqslant \frac{n_2W_n - 2^{-1}(n_2+1)u}{(12^{-1}u^2(n_2-1))^{1/2}} \leqslant b\,|\,\xi_{n_2} = u\right) - (\Phi(b) - \Phi(a))\right| \leqslant A/\sqrt{n}.$$

由此式即得

$$P(B_n|u) = \Phi\left(\frac{(n_2/\sqrt{n} - u/2)\sqrt{12}}{u(n_2-1)^{1/2}}\right) - \Phi\left(\frac{-(n_2/\sqrt{n} + u/2)\sqrt{12}}{u(n_2-1)^{1/2}}\right) + O\left(\frac{1}{\sqrt{n}}\right). \tag{1.41}$$

但对 $u\in(0,1)$一致地有

$$\Phi\left(\frac{(n_2/\sqrt{n} - u/2)\sqrt{12}}{u(n_2-1)^{1/2}}\right) - \Phi\left(\frac{-(n_2/\sqrt{n} + u/2)\sqrt{12}}{u(n_2-1)^{1/2}}\right) \geqslant \Phi(\sqrt{6}) - \Phi(-\sqrt{6}) + O\left(\frac{1}{\sqrt{n}}\right). \tag{1.42}$$

根据(1.41)和(1.42)，得

$$\lim\inf P(B_n) \geqslant \Phi(\sqrt{6}) - \Phi(-\sqrt{6}). \tag{1.43}$$

由(1.40)和(1.43)，得

$$\begin{aligned}\lim\inf P(A_{1n}\cap A_{2n}\cap B_n) &\geqslant \lim\inf(P(A_{1n}) + P(A_{2n}) + P(B_n)) \\ &\quad - 2 \geqslant \lim\inf P(A_{1n}) \\ &\quad + \lim\inf P(A_{2n}) + \lim\inf P(B_n) - 2\end{aligned}$$

$$\geqslant 2(\Phi(2) - \Phi(-2)) + \Phi(\sqrt{6}) - \Phi(-\sqrt{6}) - 2 \equiv q > 0.$$

这证明了(1.39).

定义事件

$$E_n = \{方程(1.35)无解\}.$$

往下证明存在常数 $r>0$,使当 n 充分大时,对于 $t_n \in L_{n1}$, $u_n \in L_{n2}$ 和 $w_n \in L_{n3}$,一致地有下式成立:

$$d_n \equiv P(E_n | \xi_{n_1} = t_n, \xi_{n_2} = u_n, W_n = w_n) \geqslant r. \qquad (1.44)$$

在此注意,在(1.44)式右端标出的条件下,V_n 的条件分布就等于 $n-n_2$ 个 iid. $R(u_n,1)$变量之和的分布. 因为

$$u_n \in L_{n2} \Rightarrow 1/3 \leqslant 1 - u_n \leqslant 1,$$

通过初等计算,或引用局部极限定理(见文献[49]中第7章)易证,若用 $f(t,u_n)$记随机变量

$$\begin{aligned} Q_n &\equiv Q_n(V_n) \\ &= 2\sqrt{3}(1-u_n)^{-1}(n-n_2)^{-1/2}(V_n - (1+u_n)(n-n_2)/2) \end{aligned}$$

在条件$\{\xi_{n_1}=t_n, \xi_{n_2}=u_n, W_n=w_n\}$之下的条件密度,则当 $n\to\infty$时,有

$$\sup\{|f(t,u_n) - \varphi(t)| : u_n \in L_{n2}, t \in R^1\} \to 0, \qquad (1.45)$$

此处 $\varphi(t)=(2\pi)^{-1/2}\exp(-t^2/2)$.

现将方程(1.36)写为

$$\sum_{i=1}^{n}\operatorname{sgn}(\xi_i - \alpha) + n_2 W_n + V_n - n\alpha = 0, \qquad (1.46)$$

对任一实数 α,恰有一个 V_n 值使(1.46)成立,把这个值记为 $V_n(\alpha)$,显然,当 α 在开区间(ξ_i, ξ_{i+1})内增加时,$V_n(\alpha)$线性地增加,而当 α 由 ξ_i-0变到 ξ_i+0时,$V_n(\alpha)$有跳跃2,由此可知,若

$$V_n \in (V_n(\xi_i - 0), V_n(\xi_i + 0)) - \{V_n(\xi_i)\}, \qquad (1.47)$$

则方程(1.46)无解. (1.47)中的区间长度为2,且对不同的 i,所得的区间没有公共点. 当 α 由 ξ_{n_1}上升到 ξ_{n_2}时,α 至少穿过$2\sqrt{n}-3$个 ξ_i. 记

$$D_n=\bigcup_{i=n_1}^{n_2}[(V_n(\xi_i-0),V_n(\xi_i+0))-\{V_n(\xi_i)\}],$$

则 D_n 的 Lebesgue 测度 $|D_n|\geqslant 2(2\sqrt{n}-3)$. 下证：在

$$\xi_{n_1}\in L_{n1},\xi_{n_2}\in L_{n2},\text{和 } W_n\in L_{n3}\tag{1.48}$$

的条件下，有

$$D_n\subset[3n/8-21\sqrt{n}/4-5,3n/8+9\sqrt{n}/4].\tag{1.49}$$

事实上，在(1.48)成立时，D_n 中任一点不小于

$$\begin{aligned}&(n\alpha-(n/2+\sqrt{n})W_n\\&\quad-\sum_{i=1}^{n}\operatorname{sgn}(\xi_i-\xi_{n_1}))|_{\alpha=1/2-2/\sqrt{n},W_n=1/4+2/\sqrt{n}}\\&=n(1/2-2/\sqrt{n})-(n/2+\sqrt{n})(1/4+2/\sqrt{n})\\&\quad-(n-2[n/2-\sqrt{n}]+1)\\&\geqslant 3n/8-21\sqrt{n}/4-5.\end{aligned}$$

类似地，在条件(1.48)之下，D_n 中任一点不大于

$$\begin{aligned}(n\alpha-(n/2+\sqrt{n})W_n-\sum_{i=1}^{n}\operatorname{sgn}(\xi_i-\xi_{n_2}))|_{\alpha=\frac{1}{2}+\frac{2}{\sqrt{n}},W_n=1/4-1/\sqrt{n}}&\\=n/2-(n/2+\sqrt{n})(1/4-1/\sqrt{n})&\\-(n-2[n/2+\sqrt{n}]+1)&\\\leqslant 3n/8+9\sqrt{n}/4.&\end{aligned}$$

这证明了(1.49).

现在计算在条件(1.48)及 $V_n\in D_n$ 之下 $|Q_n|$ 的最大值. 对充分大的 n，有

$$|Q_n|\leqslant 2\sqrt{3}(1/3)^{-1}(n/2-\sqrt{n})^{-1/2}6\sqrt{n}\leqslant 90.\tag{1.50}$$

记 $C=\{Q_n(V_n):V_n\in D_n\}$，以 $|C|$ 记 C 的 Lebesgue 测度. 由 $|D_n|\geqslant 2(2\sqrt{n}-3)\geqslant 3\sqrt{n}$，当 n 充分大时，知

$$|C|\geqslant(n-n_2)^{-1/2}|D_n|\geqslant 3.\tag{1.51}$$

总结以上讨论，可知当 n 充分大时，在条件(1.48)之下，一致地有

(d_n 见(1.44))：

$$d_n \geqslant \int_C f(t,u_n)\mathrm{d}t \geqslant \int_C \varphi(t)\mathrm{d}t - \int_C |f(t,u_n) - \varphi(t)|\mathrm{d}t. \tag{1.52}$$

根据(1.45)，上式右端第二项当 $n\to\infty$ 时，收敛于0. 至于第一项，由(1.50)知 $C\subset(-90,90)$，故由(1.51)得

$$\int_C \varphi(t)\mathrm{d}t \geqslant \int_{89\leqslant|t|\leqslant 90} \varphi(t)\mathrm{d}t = 2h > 0. \tag{1.53}$$

由(1.52)和(1.53)知，当 n 充分大时，对适合条件 $t_n\in L_{n1}$，$u_n\in L_{n2}$ 和 $w_n\in L_{n3}$ 的 (t_n,u_n,w_n) 一致地有

$$d_n \geqslant h \tag{1.54}$$

成立. 根据(1.38)，有

$$A_1\cap A_2\cap B\subset\{\xi_{n_1}\in L_{n1},\xi_{n_2}\in L_{n2},W_n\in L_{n3}\}. \tag{1.55}$$

最后，由(1.39)、(1.54)和(1.55)，得到

$$\liminf P(E_n)\geqslant hq>0.$$

于是完成了(1.34)的证明. ▌

第 2 章

M 估计的弱相合性

考虑线性模型(1.1).设 $c(\beta_0)$ 为 β_0 的一个已知函数,例如 β_0 本身,β_0 的一个指定分量或 β_0 的一个线性函数 $a'\beta_0$ 等.其次,设 $T_n \equiv T_n(Y_1,\cdots,Y_n)$ 为 $c(\beta_0)$的一个估计.若在某种意义上,当样本量 $n\to\infty$时,T_n 收敛于被估计的 $c(\beta_0)$,则称 T_n 为 $c(\beta_0)$的相合估计,或说 T_n 有相合性.在统计学中,常考虑的收敛意义有依概率收敛(记为 $T_n \xrightarrow{P} c(\beta_0)$或 $T_n \longrightarrow c(\beta_0)$in pr.)、以概率 1 收敛或称几乎必然收敛(记为 $T_n\to c(\beta_0)$a.s.)和 r 阶矩收敛(记为 $T_n \xrightarrow{M_r} c(\beta_0)$,即 $E\|T_n-c(\beta_0)\|^r\to 0$)等,在这些意义下的相合性分别称为弱相合、强相合及 r 阶矩相合.如所周知,由后两种相合性可推出弱相合性,但后两者之间无蕴含关系.由于弱相合是最基本类型的相合性,在本书中常把弱相合简称为相合.

要注意的是:β_0 未知,而只知道它属于 R^p 的某一个子集 B.随机误差列$(e_1,e_2,\cdots)$的分布一般也不知道,而只知道它属于某分布类 $\mathscr{F}$.因此,无论在什么意义下说 T_n 为 $c(\beta_0)$的相合估计,都是指收敛性 $T_n\to c(\beta_0)$要对一切 $\beta_0\in B$ 及 $(e_1,e_2,\cdots)$为 $\mathscr{F}$ 中的任一分布时都成立才行.

本章及下一章的主题是讨论 β_0 的 M 估计 $\hat{\beta}_n$ 的相合性问题.由于内容较多且方法上有差异,故把弱相合与强相合分别在本章及下一章讨论.

先证明一个简单事实:若对某一个特定的误差列$(e_1,e_2,\cdots)$,

$\hat{\beta}_n\equiv\hat{\beta}_n(Y_1,\cdots,Y_n;x_1,\cdots,x_n)$对某个$\beta_0$有$\hat{\beta}_n\to\beta_0$(在某一意义下，下同). 则对$R^p$中任一点$\beta_1$，当真参数值为$\beta_1$时，必有$\hat{\beta}_n\to\beta_1$.

事实上，把$H_\rho(\beta)$(见(1.4)式)写为

$$H_\rho(\beta)=\sum_{i=1}^n\rho[(Y_i-x_i'(\beta_1-\beta_0))-x_i'(\beta-(\beta_1-\beta_0))],$$

按M估计的定义，即得

$$\begin{aligned}&\hat{\beta}_n(Y_1,\cdots,Y_n;x_1,\cdots,x_n)\\&\quad=\hat{\beta}_n(Y_1-x_1'(\beta_1-\beta_0),\cdots,Y_n-x_n'(\beta_1-\beta_0);x_1,\cdots,x_n)\\&\qquad+(\beta_1-\beta_0),\end{aligned}$$

如果参数真值为β_1，则$Y_i-x_i'(\beta_1-\beta_0)=e_i+x_i'\beta_0$. 因此，

$$\hat{\beta}_n(Y_1-x_1'(\beta_1-\beta_0),\cdots,Y_n-x_n'(\beta_1-\beta_0);x_1,\cdots,x_n)$$

就是在真参数值为β_0时的M估计. 按假定，它当$n\to\infty$时收敛于β_0. 由此及上式即知

$$\hat{\beta}_n(Y_1,\cdots,Y_n;x_1,\cdots,x_n)\text{当}n\to\infty\text{时收敛于}\beta_1.$$

从而证明了上述论断.

这个事实表明：当我们讨论β_0的M估计$\hat{\beta}_n$的相合性时，总可以把β_0指定为某一特殊值，例如$\beta_0=0$. 这一点不要被误会为参数空间是无关紧要的，因为在M估计的定义中涉及到它. 在(1.5)中，若参数空间为B，则该式中的$\beta\in R^p$要改为$\beta\in B$，当B不同时，结果可以不同. 虽然如此，从相合性的定义中不难明白：若B_1、B_2为开集，且$B_1\subset B_2$，则当以B_2为参数空间的M估计为相合时，以B_1为参数空间的M估计也必相合. 特别是，当以R^p为参数空间的M估计为相合时，对任何开集$B\subset R^p$，以B为参数空间的M估计必相合. 这就是在定义M估计中，常取R^p为参数空间的原因.

但对$(e_1,e_2,\cdots)$的分布则不同，对$\mathscr{F}$中某些分布T_n收敛于β_0，并不保证对$\mathscr{F}$中其他的分布也如此. 只要在$\mathscr{F}$中能找到一个分布F，使当$(e_1,e_2,\cdots)$有分布F时，$T_n\to\beta_0$不成立，则T_n就没有相合性. 在考察相合性的必要条件时，这一点应牢记在心.

§ 2.1　基本定理及有关问题

估计的相合性,是大样本理论中首位的、讨论最多、最受重视的一个问题,其原因大致有两条:首先,有相合性这一点虽不能说明这个估计怎么好,但一个估计若没有相合性,则总是不好的,因为如果不论样本量多么大,估计值与真值仍有较大的可能出现显著的偏差,则很难相信在样本量不甚大(一般应用中多是如此)时,该估计会有良好表现.故通常在提出一个估计量时,总是要把有无相合性作为一个考察的对象.其次,与其他深层的大样本性质相比,相合性的要求最低,更有可能在较弱的条件下,获得深入的结果.但是,也正因为如此,可提出的问题也就更多,其中有的难度很大,至今未能彻底解决,甚至进展极慢.

关于 M 估计的相合性问题,除了 LSE 这个特例研究得较早(始于 60 年代)并得出较完整的结果以外,对另一个重要特例 LADE 研究得也较多,但起步较晚.至于一般的 M 估计的相合性,Huber 在其 1973 年的工作中已有研究,以后陆续发表了一些论文(例如文献[66]等),都涉及这个问题.但综观在这些工作中,所提的条件过于繁多,因而还只能认为是一个初步的研究.直到 1993 年,赵林城等发表了文献[69],对 ρ 为凸时 M 估计的相合性提出了一个形式简单且很弱的充分条件.本节的目的是介绍这个基本结果,并讨论若干有关的问题,因其证明过于复杂,为不打断论述,故放到本章 § 2.2 中.

一、基本定理的表述

考虑线性模型(1.1),设 ρ 为 R^1 上的非单调的凸函数,则按(1.5)式定义的 β_0 的 M 估计 $\hat{\beta}_n$ 存在.仍如前,以 Ψ_- 和 Ψ_+ 分别记 ρ 的左、右导数.

定理 2.1　设在模型(1.1)中随机误差 $e_1, e_2, \cdots$ 为 iid.,又存在函数 Ψ,满足(1.17)以及以下诸条件:

1° $E\Psi(e_1)=0$.

2° 存在常数 $c_0>0,\Delta>0$，使

$$|E\Psi(e_1+u)|\geqslant c_0|u|,\ |u|<\Delta. \tag{2.1}$$

3° 存在常数 $\Delta>0$，使

$$E\Psi^2(e_1\pm\Delta)<c_1<\infty. \tag{2.2}$$

记 $S_n=x_1x_1'+\cdots+x_nx_n'$，则当

$$\lim S_n^{-1}=0 \tag{2.3}$$

时，$\hat{\beta}_n$ 为 β_0 的相合估计.

注 2.1 条件(2.3)隐含了 S_n 当 n 充分大时满秩的要求，这相当于要求矩阵$(x_1\cdots x_n)$有秩 p，这是必要的，否则，β_0 为不可估计.

注 2.2 条件(2.1)包含了当 $0<|u|<\Delta$ 时，$E\Psi(e_1+u)\neq 0$. 把这与 1°结合，并注意到 Ψ 的非降性，即知 $E\Psi(e_1+u)>0$ （当 $u>0$）；$E\Psi(e_1+u)<0$ （当 $u<0$）. 根据定理 1.4，这保证了条件(1.9)满足. 于此可以看到条件 1°、2°的统计意义之所在. 但这里还有一个问题：定理 1.4 中的条件(1.22)要求对一切 $t\neq 0$ 满足，而本定理的条件 2°只要求(1.22)当 $0<|t|<\Delta$ 时满足. 对这问题解释如下：若(1.22)只在一个有限的区间 $|t|<\Delta$ 满足，则可以证明：由(1.19)定义的函数 g 也在 $|t|<\Delta$ 内有意义，而在 $|t|\geqslant\Delta$ 可以无意义或为 $\pm\infty$. 这时，只要条件(1.20)满足，仍可证明：在 g 有意义的范围内，它以 0 点为其唯一最小值点.

条件(2.1)比为了保证满足(1.9)所需要的条件多一些，其直观意义可解释如下：暂设

$$h(t)=E\rho(e_1+t)=\int_{-\infty}^{\infty}\rho(u+t)\mathrm{d}F(u)$$

存在（F 为 e_1 的分布）且可对 t 在积分号下两次求导，则有

$$h''(0)=\int_{-\infty}^{\infty}\Psi'(u)\mathrm{d}F(u).$$

如果(2.1)满足，则 $\Psi'(u)\geqslant c$ （当 $|u|<\Delta$），又因 Ψ 非降，Ψ' 处处非负，故只要知道 $P(|e_1|<\Delta)>0$，就有 $h''(0)>0$. 因而当 $|t|\to 0$ 时，$h(t)-h(0)$ 只是二阶而非更高阶的无穷小. 当 $h(t)-h(0)$ 有更

高阶的无穷小时，就模糊了 $t=0$ 是 $h(t)$ 的唯一最小点的特性，而这一特性是相合性得以成立的关键所在. 这样，就给了这个统计意义不甚明确的条件以一个直观的解释了. 当然，这不能代替严格的数学分析. 在后面我们还要讲到这个问题.

注 2.3　条件(2.2)与(1.9)满足与否完全无关. 在 $\rho(u)=u^2$ 即 LSE 的场合，这一条件相当于要求 e_1 的方差存在. 在早期关于大数定律的论证中，要求方差有限. 但后来证明了至少在 iid. 的场合，这一条件实属多余. 因此，也可以提出问题：条件(2.2)可否去掉或减轻呢？我们将在 § 2.3 第一段中回答这个问题.

在已发表的文献中，与定理 2.1 最接近的结果是文献[67]的定理 2.1. 其中考虑线性模型(1.1)，并用(1.30)的解 $\hat{\beta}_n$ 作为 β_0 的 M 估计，其中关于 $\{x_i\}$ 的条件就是(2.3). 对 $\{e_i\}$ 和 Ψ 的要求是：

(1) $e_1, e_2, \cdots$ 为 iid.；

(2) Ψ 非降；

(3) $E\Psi(e_1)=0$；

(4) $E\Psi^2(e_1)<\infty$；

(5) 存在常数 $c>0$、$d>0$，使

$$P(|e_1|\leqslant c)>0,$$

$$\inf\{z^{-1}(\Psi(u+z)-\Psi(u)):0<|z|<c,|u|<d\}>0.$$

在这些条件中，(1)～(4)在定理 2.1 的假定中也有，且此处的条件(4)比(2.2)还稍弱一点. 但是，这些条件不能保证(1.30)有解. 事实上，(1.31)定义的 Ψ 适合这些条件，但(1.30)并不以概率 1 有解. 为了使(1.30)保证有解，须增加假定 Ψ 处处连续. 这等于要求在定理 2.1 中假定 ρ 处处连续可导. 因而排除了不少重要的情况，例如 LADE. 另外，此处的条件(5)对应于(2.1)，但远比(2.1)为强，例如按条件(5)，ρ 在 0 点的某一邻域内必须是严格凸的. 这一条件也排除了 LADE.

顺便指出，C. R. Rao 和本书作者之一在 1992 年发表了论文[56]，其中对由方程(1.30)所决定的 M 估计 $\hat{\beta}_n$，或更一般地，对任何满足条件：

$$\| S_n^{-1/2}\sum_{i=1}^{n}\Psi(Y_i-x_i'\beta_n)x_i\| \xrightarrow{P} 0 \quad (当 n\to\infty)$$

的 β_n，当 Ψ 非降时，β_n 的相合性可在与基本定理 2.1 同样的条件下得到. 当然，在此仍须考虑 β_n 的存在问题.

定理 2.1 可略加推广为下面的形式：

定理 2.2 考虑线性模型

$$Y_{ni} = x_{ni}'\beta_{n0} + e_i \quad (1 \leqslant i \leqslant n, n \geqslant 1), \tag{2.4}$$

设 $e_1, e_2, \cdots$ 独立，定理 2.1 的条件 1°～3°对每个 e_i 成立，且其中的常数 c 和 Δ 都可以不依赖 i 取得. 记 $S_n = x_{n1}x_{n1}' + \cdots + x_{nn}x_{nn}'$，则当(2.3)成立时，有

$$\beta_n - \beta_{n0} \xrightarrow{P} 0 \quad (当 n \to \infty), \tag{2.5}$$

此处 β_n 指在模型(2.4)中的 β_{n0}用函数 ρ 所作的 M 估计.

细察定理 2.1 的证明(见§2.2)，不难看出，只须对该证明作些微小的修改，即可证明本定理.

二、LADE 的相合性

由凸 ρ 产生的 M 估计，除 LSE 外，最重要的无疑要数 LADE，有鉴于此，值得把定理 2.1 用于这个特例所得到的结果单独列为一条定理：

定理 2.3 设在模型(1.1)中，$e_1, e_2, \cdots$ 为 iid.，有中位数 0. 次设 e_1 的分布函数 F 满足条件：

$$|1 - 2F(u)| \geqslant c|u| \quad (当 |u| < \Delta),$$

其中 $c>0, \Delta>0$ 都是常数. 则当(2.3)式成立时，β_0 的 LADE β_n 是相合的.

为了证明此定理，只须取 $\alpha\in[-1,1]$，定义

$$\Psi(u) = \begin{cases} \mathrm{sgn}(u), & u\neq 0; \\ \alpha, & u=0, \end{cases}$$

使 $E\Psi(e_1)=0$. 这种 α 的存在由 $\mathrm{med}(e_1)=0$得出. 不难看出：在本定理对 F 的假定之下，当这样选择 α 时，所定义的 Ψ 满足定理2.1 的条件1°～3°.

这个结果确实可视为关于 LADE 相合性研究中的一个突出

成就，因为自70年代以来，有不少统计学者都在这个问题上下过功夫，发表了一些结果，但是其中个别的在证明中有不可补救的漏洞，而一般的又要求条件过强，例如在 e_i 的公共分布上，有的假定在0的邻域内有非0的连续导数，而本定理甚至不须 F 在0点连续。对于$\{x_i\}$，有的假定$\{x_i\}$有界，当 $n\to\infty$时，S_n/n 收敛于一个正定方阵等等。直到1990年，包括本书作者在内的几位作者发表了论文[4]，才对这一问题作了比较实质性的改进。他们在$\max\limits_{1\leqslant i\leqslant n}x_i'S_n^{-1}x_i\to 0$的条件下证明了 LADE 的相合性(顺便指出：虽然在文献[4]中也假定了公共分布 F 在0点附近有连续导数，但这是为了证明 LADE 的渐近正态性(见第4章)而设的，对相合性并无必要)。但此条件仍比(2.3)强。1991年 Pollard 在文献[50]中基本上重复了文献[4]中的工作。

三、线性函数 M 估计的相合性

设 $c\neq 0$为 p 维向量，$c'\beta_0$的 M 估计定义为 $c'\hat{\beta}_n$，在定理2.1的条件下，$\hat{\beta}_n$ 为相合，故 $c'\hat{\beta}_n$ 当然相合。但由于条件(2.3)保证了一切线性估计都相合，其要求可能过多了，事实上的确如此。

定理2.4　设在模型(1.1)之下 $e_1,e_2,\cdots$为 iid.，以 $\hat{\beta}_n$ 记 β_0的用凸 ρ 作出的 M 估计，且满足定理2.1的条件1°～3°。则当

$$\lim c'S_n^{-1}c = 0 \tag{2.6}$$

时，$c'\hat{\beta}_n$ 为 $c'\beta_0$的相合估计。

在 LSE 的情况，这个结果是周知的，且容易证明，它根据 $c'\hat{\beta}_n$ 无偏，且 $\mathrm{Var}(c'\hat{\beta}_n)=\mathrm{Var}(e_1)c'S_n^{-1}c\to 0$(这个证明只须$\{e_i\}$满足 Gauss-Markov 条件即可，不须 iid.)。在一般 M 估计的情况，证明也不难，但不明显。以下进行证明：记

$$S_n^{-1}=\begin{pmatrix} * & * \\ * & B_n \end{pmatrix},$$

其中 B_n 为 $p-1$阶方阵。找常数 $h_n^2>0$，使 $h_n^2B_n\to 0$。设 $c=(c_1,\cdots,c_p)'$的第一元 $c_1\neq 0$，作 p 阶方阵

$$Q_n = \begin{pmatrix} c_1 c_2 \cdots c_p \\ 0 \\ \vdots \; h_n I_{p-1} \\ 0 \end{pmatrix},$$

这里 I_{p-1} 为 $p-1$ 阶单位阵. 令

$$\beta_{n0} = Q_n \beta_0,\ x_{ni} = Q_n'^{-1} x_i\ \ (1 \leqslant i \leqslant n),$$

而把模型(1.1)写为(2.4)的形状,记 $\tilde{S}_n = \sum_{i=1}^{n} x_{ni} x_{ni}'$,则

$$\tilde{S}_n = Q_n'^{-1} S_n Q_n^{-1}, \tilde{S}_n^{-1} = Q_n S_n^{-1} Q_n',$$

按 Q_n 的定义,易见

$$\tilde{S}_n^{-1} = \begin{pmatrix} c' S_n^{-1} c & * \\ * & h_n^2 B_n \end{pmatrix}.$$

由 $h_n^2 B_n \to 0$ 及(2.6)知 $\tilde{S}_n^{-1} \to 0$,故依定理2.2,应有 $\hat{\beta}_n - \beta_{n0} \xrightarrow{P} 0$ 成立,$\hat{\beta}_n$ 为(2.4)之下 β_0 的用 ρ 作出的 M 估计. 易见 β_{n0} 的第一元为 $c'\beta_0$,而 $\tilde{\beta}_n = Q_n \hat{\beta}_n$,其第一元为 $c'\hat{\beta}_n$. 故得 $c'\hat{\beta}_n \xrightarrow{P} c'\beta_0$. 本定理得证. ▎

四、自变量为随机的情况

设有 p 维自变量 X 和一维因变量 Y,以 $Y|X$ 记给定 X 时,Y 的条件分布. 假定此条件分布的某一个特征是 X 的线性函数 $\alpha_0 + X'\beta_0$,记

$$e \equiv e(X,Y) = Y - (\alpha_0 + X'\beta_0)$$

有模型 $Y = \alpha_0 + X'\beta_0 + e$,对 (X,Y) 进行 n 次独立观察,得到样本 $(X_1,Y_1),\cdots,(X_n,Y_n)$,在形式上得到一个与(1.2)一样的模型:

$$Y_i = \alpha_0 + X_i'\beta_0 + e_i\ \ (1 \leqslant i \leqslant n, n \geqslant 1; e_i = e(X_i,Y_i)), \tag{2.7}$$

$(X_1,Y_1),\cdots,(X_n,Y_n)$ 为 iid.,其公共分布就是 (X,Y) 的分布. 选凸函数 ρ 用以作模型(2.7)中 α_0 与 β_0 的 M 估计 $\hat{\alpha}_n$、$\hat{\beta}_n$. 关于它们的相合性,根据定理2.2,容易证明以下的结果.

定理2.5 以 ξ_X 记一随机变量,其分布与 $e(X,Y)|X$ 相同,设

满足以下两条件：

1° ξ_X 满足定理2.1的条件1°～3°，其中的常数与 X 无关.

2° X 的分布不退化到任一 $p-1$维超平面上. 则 $\hat{\alpha}_n$、$\hat{\beta}_n$ 分别是 α_0、β_0的相合估计.

证 由定理的假设，根据定理2.2，只须证明(2.3)以概率1成立，其中 $S_n=W_1W_1'+\cdots+W_nW_n'$，而 $W_i'=(1,X_i')$. 然后在而定 $X_1,X_2,\cdots$，这一条件下去考虑模型(2.7)即可.

为证(2.3)，记

$$W'=(1,X'),$$

把如上定义的 W 视为一个 $p+1$维随机向量，则其支撑 K 不能全落在 R^{p+1}中一个过原点的超平面上. 因为若不然，则存在 $\alpha\in R^1$和 $\beta\in R^p$，使 α、β 不全为0，且

$$P(\alpha+X'\beta=0)=1.$$

这与假定2°矛盾.

以 T_r 记 R^{p+1}中以原点为中心，r 为半径的球. 由上述知，当 r 充分大时，$K\cap T_r$ 不能落在 R^{p+1}中任一过原点的 p 维超平面上. 因此，若定义

$$\tilde{W}=W/\max(1,\|W\|/r),$$

则 $\tilde{W}$ 的分布不会退化到 R^{p+1}中任一过原点的 p 维超平面上，再加上 $\tilde{W}$ 有界，即知 $E(\tilde{W}\tilde{W}')$存在，且为正定的. 因此，若记

$$\tilde{W}_i=W_i/\max(1,\|W_i\|/r),\quad \tilde{S}_n=\sum_{i=1}^{n}\tilde{W}_i\tilde{W}_i',$$

则根据强大数定律，有 $\tilde{S}_n/n\to E(\tilde{W}\tilde{W}')>0$，a.s.. 由此推出

$$\tilde{S}_n^{-1}\to 0,\text{a.s.}. \tag{2.8}$$

但由 $\tilde{W}_i$ 的定义知，$\tilde{W}_i\tilde{W}_i'\leqslant W_iW_i'$，因而 $S_n\geqslant\tilde{S}_n$. 故 $\lambda(S_n)\geqslant\lambda(\tilde{S}_n)$，将此与(2.8)结合，即得(2.3). 从而定理得证. ▌

定理的假定2°是必要的. 否则，β_0为不可估.

本定理条件的优点在于没有对 X、Y(或其函数)的矩提出要求. 但在条件1°中"与 X 无关"的要求不好验证. 一个在实用上可

以想象的情况如下：因为变量 Y 由两部分构成：特征 $\alpha_0+X'\beta$ 及误差 e，后者与 X 独立，且其分布不依赖于 X. 这时，定理的条件等于要求 e 满足定理2.1的1°～3°.

五、模型含常数项时

设有线性模型(1.2)，由于(1.2)只是(1.1)的一个特例，故 α_0、β_0的 M 估计 $\hat{\alpha}_n$ 和 $\hat{\beta}_n$ 的弱相合问题也适用定理2.1～2.4. 这里我们要证明：对这一模型条件(2.3)可表为一种等价形式，可能有其方便之处.

对模型(1.2)，有

$$S_n=\sum_{i=1}^{n}\begin{pmatrix}1\\x_i\end{pmatrix}(1,x_i')=\begin{pmatrix}n & n\bar{x}_n'\\ n\bar{x}_n & S_{n0}\end{pmatrix},\quad S_{n0}=\sum_{i=1}^{n}x_ix_i',$$

这里 $\bar{x}_n=\sum_{i=1}^{n}X_i/n$. 若记 $T_n=\sum_{i=1}^{n}(x_i-\bar{x}_n)(x_i-\bar{x}_n)'$，则根据“四块求逆”的公式，可得

$$S_n^{-1}=\begin{pmatrix}(n-n^2\bar{x}_n'S_{n0}^{-1}\bar{x}_n)^{-1} & *\\ * & T_n^{-1}\end{pmatrix}.$$

因此，条件(2.3)等价于以下两式同时成立：

$$\lim T_n^{-1}=0, \tag{2.9}$$

$$\lim(n-n^2\bar{x}_n'S_{n0}^{-1}\bar{x}_n)^{-1}=0. \tag{2.10}$$

有趣的是：(2.10)是(2.9)的后果，因此根据定理2.1，在其中的条件1°～3°满足的前提下，由(2.9)成立可推出 $\hat{\alpha}_n$、$\hat{\beta}_n$ 都是弱相合. 其实，$\hat{\beta}_n$ 的弱相合可由定理2.4得出；但 $\hat{\alpha}_n$ 的相合性，则有赖于由(2.9)推出(2.10).

为了证明，注意：因 $S_{n0}=T_n+n\bar{x}_n\bar{x}_n'$，有(见文献[54]，p. 33)：

$$S_{n0}^{-1}=T_n^{-1}-nT_n^{-1}\bar{x}_n\bar{x}_n'T_n^{-1}/(1+n\bar{x}_n'T_n^{-1}\bar{x}_n).$$

由此得出

$$(n-n^2\bar{x}_n'S_{n0}^{-1}\bar{x}_n)^{-1}=n^{-1}+\bar{x}_n'T_n^{-1}\bar{x}_n.$$

因此，要证明(2.10)，只须证明

$$\lim\bar{x}_n'T_n^{-1}\bar{x}_n=0. \tag{2.11}$$

为证上式，先证以下的预备事实：设 G 是一个由某些无穷实数列 $a=(a_1,a_2,\cdots)$ 组成的集. 记 $\bar{a}_n=\sum_{i=1}^{n}a_i/n$，$\sigma_n^2(a)=\sum_{i=1}^{n}(a_i-\bar{a}_n)^2$，设 G 满足以下条件：

(1) $\sup\{|a_i|:a\in G\}<\infty\ (i=1,2,\cdots)$；

(2) $\inf\{\sigma_n^2(a):a\in G\}\to\infty$ （当 $n\to\infty$）.

则对 G 中的 a 一致地有下式成立：

$$\bar{a}_n^2/\sigma_n^2(a)\to 0. \tag{2.12}$$

证明如下：任给 $\varepsilon>0$，求自然数 m，使$4/m\leqslant\varepsilon$. 其次找充分大的 n_0，使 $\sigma_{n_0}(a)\sqrt{\varepsilon}>2\sup\{|a_i|:a\in G,1\leqslant i\leqslant m\}$. 对任取的 $a\in G$，用 $n\geqslant\max(n_0,m)$ 计算 $\bar{a}_n^2/\sigma_n^2(a)$，分两种情况：若 $\bar{a}_n^2\leqslant\sigma_n^2(a)\varepsilon$，则 $\bar{a}_n^2/\sigma_n^2(a)\leqslant\varepsilon$；若 $|\bar{a}_n|>\sigma_n(a)\sqrt{\varepsilon}$，则根据 n_0的取法及当 $n\geqslant n_0$时，$\sigma_n(a)\geqslant\sigma_{n0}(a)$，有

$$\sigma_n^2(a)\geqslant\sum_{i=1}^{m}(a_i-\bar{a}_n)^2\geqslant\sum_{i=1}^{m}(|\bar{a}_n|-|\bar{a}_n|/2)^2=m|\bar{a}_n|^2/4.$$

因此 $\bar{a}_n^2/\sigma_n^2(a)\leqslant 4/m\leqslant\varepsilon$. 由 $\varepsilon>0$的任意性及 m、n_0与 a 无关，即得(2.12)对 $a\in G$ 一致成立.

现设

$$G=\{(a'x_1,a'x_2,\cdots):\|a\|=1\},$$

以上条件(1)自然成立；条件(2)由(2.9)推出. 因为

$$\sum_{i=1}^{n}(a'x_i-a'\bar{x}_n)^2=a'T_na,$$

故其在$\|a\|=1$上的最小值(即 T_n 的最小特征根 $\lambda(T_n)$)随 $n\to\infty$而趋向于无穷.

按上述预备事实，对满足条件$\|a_n\|=1$的任一列$\{a_n\}$，有

$$(a_n'\bar{x}_n)^2/a_n'T_na_n\to 0.$$

取 $a_n=T_n^{-1}\bar{x}_n/\|T_n^{-1}\bar{x}_n\|$，由上式即得(2.11)(若 $T_n^{-1}\bar{x}_n=0$，则 $\bar{x}_n'T_n^{-1}\bar{x}_n=0$，这对(2.11)的成立是无害的).

因为 T_n 只有 p 阶，而直接在模型(1.2)中用条件(2.3)所要验证的方阵为 $p+1$阶，因此用(2.9)可能是很方便的. 这个形式的

意义是：从其更易解释条件(2.3)的统计意义. 例如，以 $p=1$(x_i 为一维)的情况而言，条件(2.9)是

$$\lim\sum_{i=1}^{n}(x_i-\bar{x}_n)^2=\infty.$$

这个条件的含义是：x_i 不能过于聚集在某点 a(如 $\bar{x}_n$)的附近. 这一要求从统计上说是很好理解的：若 x_i 过于聚集在某点 a 附近，则回归直线犹如支在$(a,\alpha_0+a\beta_0)$点处的一支杠杆. 在 a 点近傍的 x_i 处的观察值 y_i 只要稍有误差，就等于在支点附近拨动这根杆，则将引起回归线斜率的显著变动. 因此，在这种情况下，回归参数难于有较精的估计，从而也就难以达到相合的要求.

类似地，在多维的情况，(2.9)相当于要求 x_i 不能过于集中在某点(如 $\bar{x}_n$)附近，且其方向不能过于偏在一隅. 就是说，除了$\|x_i-\bar{x}_n\|$ $(i\geqslant1)$有足够的大小外，其方向(在 R^p 中单位球上的投影)不能局限在一个过小的空间角度内，因为这一情况相当于近似的共线性，使估计量的精度大受影响.

基本定理2.1讲的是 ρ 为凸的情形；对 ρ 为非凸的情况，在下一章中考虑 $\hat{\beta}_n$ 的强相合问题时有讨论. 目前在文献中，还没有见到过 ρ 为非凸时的 $\hat{\beta}_n$ 弱相合的结果.

§ 2.2　基本定理的证明

基本定理的证明所用的工具是初等的，但推理很繁，初读时可暂时略过，并不影响对后文的理解.

令

$$\beta_{n_0}=S_n^{1/2}\beta_0,\quad x_{ni}=S_n^{-1/2}x_i\quad(1\leqslant i\leqslant n),$$

而把模型(1.1)写为

$$Y_i=x_{ni}'\beta_{n_0}+e_i\quad(1\leqslant i\leqslant n,n\geqslant1).\qquad(2.13)$$

注意：在此有$\sum_{i=1}^{n}x_{ni}x_{ni}'=I_p$. 以 $\hat{\beta}_{n_0}$记模型(2.13)中的 β_{n_0}的(使用 ρ

作出的)M 估计，则易见 $\hat{\beta}_{n_0}=S_n^{1/2}\hat{\beta}_n$，$\hat{\beta}_n$ 是(1.1)中的 β_0的 M 估计. 由于 $S_n^{-1}\to 0$，为证 $\hat{\beta}_n$ 相合，只须证明

$$\hat{\beta}_{n_0}-\beta_{n_0}=O_p(1). \tag{2.14}$$

不失普遍性，不妨设 $\beta_0=0$. 这时 $\beta_{n_0}=0$，且 $Y_i=e_i$. 以 U 记 R^p 中的单位球面. 定义

$$\begin{aligned} D_n(\beta) &= \sum_{i=1}^{n}(\rho(e_i - x_{ni}'\beta) - \rho(e_i)) \\ &= \sum_{i=1}^{n} \int_0^{-x_{ni}'\beta}\Psi(e_i + t)\mathrm{d}t, \end{aligned}$$

后一等式由于 ρ 为凸，故在任意有界区间上为绝对连续的. 令

$$D(\gamma,L) = D_n(L\gamma) = \sum_{i=1}^{n} \int_0^{-Lx_{ni}'\gamma}\Psi(e_i + t)\mathrm{d}t, \tag{2.15}$$

对 $\gamma\in U$，$L>0$. 因为 $D_n(\beta)$是 β 的凸函数，按定理1.3的3°及 $\hat{\beta}_{n_0}$的定义，有

$$P(\|\hat{\beta}_{n0}\| \geqslant L) \leqslant P(\inf_{\gamma\in U} D(\gamma,L) \leqslant 0). \tag{2.16}$$

把 U 分解为两个非交集的并 $A\cup A^c$. 由 $D_n(\beta)$的凸性容易看出：若 $L_1>0$，$L_2>0$，而 $L\geqslant\max(L_1,L_2)$，且

$$\inf_{\gamma\in A} D(\gamma,L_1) > 0, \inf_{\gamma\in A^c} D(\gamma,L_2) > 0,$$

则有$\inf_{\gamma\in U} D(\gamma,L)>0$. 因此有

$$\begin{aligned} &\{\inf_{\gamma\in U} D(\gamma,L) \leqslant 0\} \\ &\quad \subset \{\inf_{\gamma\in A} D(\gamma,L_1) \leqslant 0\} \\ &\quad \cup \{\inf_{\gamma\in A^c} D(\gamma,L_2) \leqslant 0\}. \end{aligned} \tag{2.17}$$

这个关系是整个证明的出发点.

设 $\delta>0$，$\varepsilon_1>0$，$0<\eta<1$. 定义

$$J = \{j: 1 \leqslant j \leqslant n; \|x_{nj}\| > \delta\}, \quad J^c = \{1,\cdots,n\} - J, \tag{2.18}$$

$$S(\delta) = \sum\nolimits_{i\in J^c} x_{ni}x_{ni}' = \sum_{i=1}^{n} x_{ni}x_{ni}' I(\|x_{ni}\| \leqslant \delta), \tag{2.19}$$

$$A = \{\gamma: \gamma \in U, \gamma' S(\delta)\gamma \geqslant \varepsilon_1\}, \ A^c = U - A, \tag{2.20}$$

$$I_\gamma=\{j:1\leqslant j\leqslant n,|x_{nj}'\gamma|\geqslant\eta\},$$
$$I_\gamma^c=\{1,\cdots,n\}-I_\gamma. \qquad (2.21)$$

由 Ψ 非降及定理2.1的条件1°、2°，易见 $\Psi(-\infty)<0<\Psi(\infty)$. 故可找到 K，使 $\Psi(-\infty)<-K<K<\Psi(\infty)$.

给定 $\varepsilon>0$，找 ε_2，使

$$0<\varepsilon_2<2^{-8}(c_1p)^{-1/2}\varepsilon^{1/2}K. \qquad (2.22)$$

此处 c_1见(2.2). 把 U 分为 M 个非交集 $\tilde{U}_1,\cdots,\tilde{U}_M$ 的并，使每个 $\tilde{U}_i$ 的直径不超过 ε_2，再取 ε_1、ε_3，使

$$0<\varepsilon_1<\min(1/2,2^{-16}(Mc_1)^{-1}K^2\varepsilon),$$
$$0<\varepsilon_3<\min(1,(6p)^{-1}\varepsilon_1). \qquad (2.23)$$

把 U 分为 N 个非交集 $\tilde{V}_1,\cdots,\tilde{V}_N$ 的并，使每个 $\tilde{V}_i$ 的直径不超过 ε_3/p. 令

$$L_1=2^6(c_1NP)^{1/2}(c_0\varepsilon_1\varepsilon^{1/2})^{-1},\quad 0<\delta<\Delta/L_1, \qquad (2.24)$$
$$m=[p/\delta^2](\text{不超过 } p/\delta^2 \text{ 的最大整数}). \qquad (2.25)$$

再取 $a>0$充分大，使

$$P(\Psi(e_1+a)\leqslant K)<(16m)^{-1}\varepsilon,$$
$$P(\Psi(e_1-a)\geqslant -K)<(16m)^{-1}\varepsilon. \qquad (2.26)$$

最后，取 η、L_2，使

$$0<\eta<\min((2\sqrt{m})^{-1},(2^8m\sqrt{c_1})^{-1}K\varepsilon), \qquad (2.27)$$

$$L_2\geqslant\max(2a/\eta,2^7am\sqrt{c_1}(K\varepsilon)^{-1}). \qquad (2.28)$$

上面引进了许多常数，其关系错综复杂. 为了明确其间没有矛盾，小结如下：

c_0、c_1、Δ：由定理假定中给出.

K：由 $\Psi(-\infty)<-K<K<\Psi(\infty)$给定.

ε：给定，$\varepsilon>0$，ε_2：由(2.22)给定.

M：由 ε_2定出的 U 的分割数.

ε_1：由 M、K、ε 定出，ε_3：由 ε_1定出.

N：由 ε_3定出的 U 的分割数.

L_1：由 N、ε_1、ε 定出，δ：由 L_1、Δ 定出.

m：由 δ 定出.

α：由 m、ε、K 定出.

η：由 m、c_1、K、ε 定出.

L_2：由 α、η、m、c_1、K 和 ε 定出.

先考虑 $\gamma\in A$ 的情形，由(2.15)及 Ψ 的非降性，有

$$\begin{aligned} D(\gamma,L_1)=&\sum_{i=1}^{n}\int_0^{-L_1x_{ni}'\gamma}\{\Psi(e_i+t)-\Psi(e_i)\}\mathrm{d}t\\ &-L_1\sum_{i=1}^{n}x_{ni}'\gamma\Psi(e_i)\\ \geqslant&\sum\nolimits_{i\in J^c}\int_0^{-L_1x_{ni}'\gamma}(\Psi(e_i+t)-\Psi(e_i))\mathrm{d}t\\ &-L_1\left\|\sum_{i=1}^{n}x_{ni}'\Psi(e_i)\right\|. \end{aligned}\tag{2.29}$$

把集合 $\tilde{V}_i\cap A$ $(1\leqslant i\leqslant N)$之中的非空者记为 $V_1,\cdots,V_{N_1}$，由于 $L_1V_j=\{L_1\gamma:\gamma\in V_j\}$的直径小于 $L_1\varepsilon_3/p$，故可以被一个边长不超过 $L_1\varepsilon_3/p$ 的闭的 p 维超立方体 T_j 所覆盖，T_j 的直径不超过 $(p(L_1\varepsilon_3/p)^2)^{1/2}\leqslant L_1\varepsilon_3$ $(1\leqslant j\leqslant N_1)$. 在 V_j 中取定一点 γ_j $(1\leqslant j\leqslant N_1)$，固定一个 j，有三种可能的情况：

1° $-x_{ni}'\beta\geqslant0$对一切 $\beta\in T_j$. 由于 T_j 是闭的，可找到 $\beta_{ij}\in T_j$，使

$$-x_{ni}'\beta_{ij}=\inf\{-x_{ni}'\beta:\beta\in T_j\}\geqslant0.$$

2° $-x_{ni}'\beta\leqslant0$对一切 $\beta\in T_j$. 这时可找到 $\beta_{ij}\in T_j$，使

$$-x_{ni}'\beta_{ij}=\sup\{-x_{ni}'\beta:\beta\in T_j\}\leqslant0.$$

3° 其他情况. 这时可找到 $\beta_{ij}\in T_j$，使 $x_{ni}'\beta_{ij}=0$.

记

$$G(t)=E\Psi(e_1+t)\quad(t\in R^1),\tag{2.30}$$

$$\Phi(e_i,t)=\Psi(e_i+t)-\Psi(e_i)-G(t).\tag{2.31}$$

由 Ψ 的非降性，及上述 β_{ij}的选择，得

$$\inf_{\gamma\in\Lambda}\sum\nolimits_{i\in J^c}\int_0^{-L_1x'_{ni}\gamma}(\Psi(e_i+t)-\Psi(e_i))\mathrm{d}t$$

$$\geqslant\inf_{1\leqslant j\leqslant N_1}\sum\nolimits_{i\in J^c}\int_0^{-x'_{ni}\beta_{ij}}(\Psi(e_i+t)-\Psi(e_i))\mathrm{d}t$$

$$\geqslant\inf_{1\leqslant j\leqslant N_1}\sum\nolimits_{i\in J^c}\int_0^{-x'_{ni}\beta_{ij}}G(t)\mathrm{d}t$$

$$-\sup_{1\leqslant j\leqslant N_1}\left|\sum\nolimits_{i\in J^c}\int_0^{-x'_{ni}\beta_{ij}}\Phi(e_i,t)\mathrm{d}t\right|. \qquad (2.32)$$

由于$\|\beta_{ij}-L_1\gamma_j\|<L_1\varepsilon_3$，由(2.23)并注意$\|\gamma_j\|=1$，有

$$\left|\sum\nolimits_{i\in J^c}(x'_{ni}\beta_{ij})^2-\sum\nolimits_{i\in J^c}(L_1x'_{ni}\gamma_j)^2\right|$$

$$\leqslant L_1^2(2+\varepsilon_3)\varepsilon_3\sum_{i=1}^{n}\|x_{ni}\|^2$$

$$=L_1^2(2+\varepsilon_3)\varepsilon_3 p<3L_1^2\varepsilon_3 p<L_1^2\varepsilon_1/2. \qquad (2.33)$$

由于β_{ij}的选择，有$|-x'_{ni}\beta_{ij}|\leqslant|L_1x'_{ni}\gamma_j|$. 再由(2.18)及(2.24)，得

$$|-x'_{ni}\beta_{ij}|\leqslant|L_1x'_{ni}\gamma_j|\leqslant L_1\delta<\Delta, 当\ i\in J^c, 1\leqslant j\leqslant N_1. \qquad (2.34)$$

由(2.1)，有

$$\int_0^{-x'_{ni}\beta_{ij}}G(t)dt\geqslant c_0(x'_{ni}\beta_{ij})^2/2.$$

再利用(2.34)，并注意由(2.19)、(2.20)以及$\gamma_j\in\Lambda$所得出的

$$\sum\nolimits_{i\in J^c}(x'_{ni}\gamma_j)^2=\gamma'_jS(\delta)\gamma_j\geqslant\varepsilon_1,$$

即可推出

$$\inf_{1\leqslant j\leqslant N_1}\sum\nolimits_{i\in J^c}\int_0^{-x'_{ni}\beta_{ij}}G(t)dt\geqslant c_0\inf_{1\leqslant j\leqslant N_1}\sum\nolimits_{i\in J^c}(x'_{ni}\beta_{ij})^2/2$$

$$\geqslant 2^{-1}c_0L_1^2\inf_{1\leqslant j\leqslant N_1}\left(\sum\nolimits_{i\in J^c}(x'_{ni}\gamma_j)^2-\varepsilon_1/2\right)$$

$$\geqslant c_0L_1^2\varepsilon_1/4. \qquad (2.35)$$

由$\|\beta_{i1}-L_1\gamma_j\|<L_1\varepsilon_3$知

$$\|\beta_{ij}\|\leqslant\|L_1\gamma_j\|+L_1\varepsilon_3=L_1+L_1\varepsilon_3<2L_1,$$

故
$$\sum_{i\in J^c}(x_{ni}'\beta_{ij})^2<4L_1^2\sum_{i=1}^n\|x_{ni}\|^2=4pL_1^2.$$
注意到 $E\Phi^2(e_i,t)\leqslant E(\Psi(e_i+t)-\Psi(e_i))^2$，有
$$E\left(\int_0^{-x_{ni}'\beta_{ij}}\Phi(e_i,t)\mathrm{d}t\right)^2$$
$$\leqslant\left|\int_0^{-x_{ni}'\beta_{ij}}\mathrm{d}t\right|\left|\int_0^{-x_{ni}'\beta_{ij}}E(\Psi(e_i+t)-\Psi(e_i))^2\mathrm{d}t\right|.$$
若 $i\in J_c$，则 $\|x_{ni}\|\leqslant\delta$，因而 $|x_{ni}'\beta_{ij}|\leqslant L_1\delta<\Delta$. 从而对区间 $[-|x_{ni}'\beta_{ij}|,|x_{ni}'\beta_{ij}|]$ 内的 t，有
$$E(\Psi(e_i+t)-\Psi(e_i))^2\leqslant 2E\Psi^2(e_i+t)+2E\Psi^2(e_i)\leqslant 4c_1,$$
式中的 c_1 见(2.2)式. 故有
$$E\left(\int_0^{-x_{ni}'\beta_{ij}}\Phi(e_i,t)\mathrm{d}t\right)^2\leqslant 4c_1(x_{ni}'\beta_{ij})^2\quad(i\in J_c).$$
由以上诸式，得到
$$P\left(\sup_{1\leqslant j\leqslant N_1}\left|\sum_{i\in J^c}\int_0^{-x_{ni}'\beta_{ij}}\Phi(e_i,t)\mathrm{d}t\right|\geqslant c_0L_1^2\varepsilon/2^3\right)$$
$$\leqslant\sum_{j=1}^{N_1}P\left(\left|\sum_{i\in J^c}\int_0^{-x_{ni}'\beta_{ij}}\Phi(e_i,t)\mathrm{d}t\right|\geqslant c_0L_1^2\varepsilon_1/2^3\right)$$
$$\leqslant\sum_{j=1}^{N_1}2^6(c_0L_1^2\varepsilon_1)^{-2}\sum_{i\in J^c}E\left(\int_0^{-x_{ni}'\beta_{ij}}\Phi(e_i,t)\mathrm{d}t\right)^2$$
$$\leqslant\sum_{j=1}^{N_1}2^8c_1(c_0L_1^2\varepsilon_1)^{-2}\sum_{i\in J^c}(x_{ni}'\beta_{ij})^2$$
$$\leqslant\sum_{j=1}^{N_1}2^{10}c_1(c_0L_1^2\varepsilon_1)^{-2}pL_1^2\leqslant 2^{10}c_1Np(c_0\varepsilon_1L_1)^{-2}=\varepsilon/4. \tag{2.36}$$
最后一个等式是根据(2.24)而得. 再利用(2.2)、(2.24)并注意到 $E\Psi(e_i)=0$ 及 $e_1,e_2,\cdots$ 独立，有
$$P\left(L_1\left\|\sum_{i=1}^n x_{ni}\Psi(e_i)\right\|\geqslant c_0L_1^2\varepsilon_1/2^3\right)$$

$$\leqslant 2^6(c_0L_1\varepsilon_1)^{-2}E\left\|\sum_{i=1}^n x_{ni}\Psi(e_i)\right\|^2 \leqslant 2^6(c_0L_1\varepsilon_1)^{-2}pc_1$$

$$= 2^{-6}N^{-1}\varepsilon < \varepsilon/4. \tag{2.37}$$

由(2.29)、(2.32)及(2.35)～(2.37)，得

$$P(\inf_{\gamma\in\Lambda} D(\gamma,L_1)\leqslant 0)<\varepsilon/2. \tag{2.38}$$

现在考虑 $\gamma\in\Lambda^c$ 的情形：由 $\varepsilon_1<1/2$，以及(2.19)和(2.20)，有

$$1/2>\varepsilon_1>\gamma' S(\delta)\gamma=\sum\nolimits_{i\in J^c}(x_{ni}'\gamma)^2$$

由此并注意到$\sum_{i=1}^n(x_{ni}'\gamma)=\gamma'\gamma=1$，有$\sum_{i\in J}(x_{ni}'\gamma)^2>1/2$. 以 $\#(J)$ 记集 J 所含元素的个数，则有

$$p=\sum_{i=1}^n\|x_{ni}\|^2\geqslant\delta^2\#(J),$$

由此知 $\#(J)\leqslant p/\delta^2$，故

$$\#(J)\leqslant[p/\delta^2]=m(\text{见}(2.25)).$$

由此及(2.21)、(2.27)，有

$$1/2<\sum\nolimits_{i\in J}(x_{ni}'\gamma)^2=\sum\nolimits_{i\in J\cap I_\gamma}(x_{ni}'\gamma)^2+\sum\nolimits_{i\in J\cap I_\gamma^c}(x_{ni}'\gamma)^2$$

$$\leqslant\sum\nolimits_{i\in J\cap I_\gamma}(x_{ni}'\gamma)^2+m\eta^2<\sum\nolimits_{i\in J\cap I_\gamma}(x_{ni}'\gamma)^2+1/4.$$

最后一个不等式根据(2.27)而得. 由此式得

$$\sum\nolimits_{i\in J\cap I_\gamma}(x_{ni}'\gamma)^2>1/4. \tag{2.39}$$

定义事件

$$A_n=\bigcap\nolimits_{i\in J}\{\Psi(e_i+\alpha)>K\}\bigcap\nolimits_{i\in J}\{\Psi(e_i-\alpha)<-K\}, \tag{2.40}$$

$$B_n=\{\alpha\sum\nolimits_{i\in J}|\Psi(e_i)|<L_2K/16\}. \tag{2.41}$$

由(2.26)、(2.28)及 $\#(J)\leqslant m$，知

$$P(A_n^c\cup B_n^c)<\varepsilon/8+P(\alpha\sum\nolimits_{i\in J}|\Psi(e_i)|\geqslant L_2K/16)$$

$$\leqslant\varepsilon/8+16\alpha(L_2K)^{-1}E\left(\sum\nolimits_{i\in J}|\Psi(e_i)|\right)$$

$$\leqslant\varepsilon/8+16\alpha m\sqrt{c_1}(L_2K)^{-1}\leqslant\varepsilon/8+\varepsilon/8=\varepsilon/4, \tag{2.42}$$

其中 $16am\sqrt{c_1}(L_2K)^{-1}\leqslant\epsilon/8$ 是根据(2.28)而得. 再根据(2.28),有

$$L_2|x'_{ni}\gamma|\geqslant L_2\eta\geqslant 2a\quad(i\in I_\gamma).$$

由此及(2.39),并注意到 Ψ 非降,即知当 $e_i\in A_n\cap B_n(i=1,\cdots,n)$ 同时成立时,有

$$\begin{aligned}
&\inf_{\gamma\in\Lambda^c}\sum\nolimits_{i\in J\cap I_\gamma}\int_0^{-L_2x'_{ni}\gamma}\Psi(e_i+t)\mathrm{d}t\\
&\geqslant\inf_{\gamma\in\Lambda^c}\sum\nolimits_{i\in J\cap I_\gamma}\left(\int_0^{a\cdot\mathrm{sgn}(-x'_{ni}\gamma)}\Psi(e_i)\mathrm{d}t+\int_{a\cdot\mathrm{sgn}(-x'_{ni}\gamma)}^{-L_2x'_{ni}\gamma}\Psi(e_i+t)\mathrm{d}t\right)\\
&\geqslant\inf_{\gamma\in\Lambda^c}\sum\nolimits_{i\in J\cap I_\gamma}K(L_2|x'_{ni}\gamma|-a)-a\sum\nolimits_{i\in J}|\Psi(e_i)|\\
&\geqslant\inf_{\gamma\in\Lambda^c}\sum\nolimits_{i\in J\cap I_\gamma}K(L_2|x'_{ni}\gamma|-L_2|x'_{ni}\gamma|/2)-L_2K/16\\
&\geqslant 2^{-1}L_2K\inf_{\gamma\in\Lambda^c}\sum\nolimits_{i\in J\cap I_\gamma}|x'_{ni}\gamma|-L_2K/16\\
&\geqslant 2^{-1}L_2K\inf_{\gamma\in\Lambda^c}\sum\nolimits_{i\in J\cap I_\gamma}(x'_{ni}\gamma)^2-L_2K/16\\
&>L_2K/8-L_2K/16=L_2K/16.
\end{aligned}\tag{2.43}$$

由(2.42)、(2.43)得

$$P\left(\inf_{\gamma\in\Lambda^c}\sum\nolimits_{i\in J\cap I_\gamma}\int_0^{-L_2x'_{ni}\gamma}\Psi(e_i+t)\mathrm{d}t\leqslant L_2K/16\right)<\epsilon/4,\tag{2.44}$$

由 Ψ 非降, $\#(J)\leqslant m$ 及 $|x'_{ni}\gamma|<\eta$ (当 $i\in I_\gamma^c$),有

$$\begin{aligned}
\int_0^{-L_2x'_{ni}\gamma}\Psi(e_i+t)\mathrm{d}t&\geqslant\int_0^{-L_2x'_{ni}\gamma}\Psi(e_i)\mathrm{d}t\\
&\geqslant-L_2\eta|\Psi(e_i)|\quad(i\in I_\gamma^c).
\end{aligned}$$

由(2.2)和(2.27)得

$$\begin{aligned}
&P\left(\inf_{\gamma\in\Lambda^c}\sum\nolimits_{i\in J\cap I_\gamma^c}\int_0^{-L_2x'_{ni}\gamma}\Psi(e_i+t)\mathrm{d}t\leqslant-L_2K/32\right)\\
&\leqslant P\left(L_2\eta\sum\nolimits_{i\in J}|\Psi(e_i)|\geqslant L_2K/32\right)
\end{aligned}$$

$$\leqslant 32\eta K^{-1}E\sum_{i\in J}|\Psi(e_i)|$$

$$\leqslant 32\eta K^{-1}\sum_{i\in J}(E\Psi^2(e_i))^{1/2}$$

$$\leqslant 32\eta K^{-1}m\sqrt{c_1}\leqslant \varepsilon/8. \tag{2.45}$$

对于 $\gamma\in\Lambda^c$，由 Ψ 的非降性，有

$$\sum_{i\in J^c}\int_0^{-L_2x_{ni}'\gamma}\Psi(e_i+t)\mathrm{d}t$$

$$\geqslant \sum_{i\in J^c}\int_0^{-L_2x_{ni}'\gamma}\Psi(e_i)\mathrm{d}t$$

$$\geqslant -L_2\sum_{i\in J^c}\Psi(e_i)x_{ni}'\gamma. \tag{2.46}$$

以 $U_1,\cdots,U_{M_1}$ 记 $\tilde{U}_1\cap\Lambda^c,\cdots,\tilde{U}_M\cap\Lambda^c$ 中的非空者，取 $\gamma_j\in U_j$ ($j=1,\cdots,M_1$)，注意到

$$\sup_{1\leqslant j\leqslant M_1}\sup_{\gamma\in U_j}\|\gamma-\gamma_j\|\leqslant\varepsilon_2,$$

由(2.2)和(2.22)，有

$$P\left(\sup_{1\leqslant j\leqslant M_1}\sup_{\gamma\in U_j}L_2\left|\sum_{i\in J^c}\Psi(e_i)x_{ni}'(\gamma-\gamma_j)\right|\geqslant L_2K/64\right)$$

$$\leqslant P\left(\left\|\sum_{i\in J^c}x_{ni}'\Psi(e_i)\right\|\geqslant K\varepsilon_2^{-1}/64\right)$$

$$\leqslant (64\varepsilon_2K^{-1})^2E\left\|\sum_{i\in J^c}x_{ni}'\Psi(e_i)\right\|^2$$

$$\leqslant (2^6\varepsilon_2K^{-1})^2c_1p<\varepsilon/16. \tag{2.47}$$

最后一不等式是根据(2.22)而得.

因为 $\gamma_j\in\Lambda^c$，有 $\gamma_jS(\delta)\gamma_j<\varepsilon_1$. 据此及(2.2)、(2.23)，有

$$P\left(\sup_{1\leqslant j\leqslant M_1}L_2\left|\sum_{i\in J^c}\Psi(e_i)x_{ni}'\gamma_j\right|\geqslant L_2K/64\right)$$

$$\leqslant\sum_{j=1}^{M_1}(64K^{-1})^2E\left|\sum_{i\in J^c}\Psi(e_i)x_{ni}'\gamma_j\right|^2$$

$$\leqslant\sum_{j=1}^{M_1}(64K^{-1})^2c_1\gamma_jS(\delta)\gamma_j\leqslant 2^{12}K^{-2}Mc_1\varepsilon_1<\varepsilon/16. \tag{2.48}$$

由(2.46)～(2.48)，得

$$P\left(\inf_{\gamma\in A^c}\sum_{i\in J^c}\int_0^{-L_2x_m'\gamma}\Psi(e_i+t)\mathrm{d}t\leqslant -L_2K/32\right)<\varepsilon/8. \tag{2.49}$$

由(2.15)、(2.44)、(2.45)和(2.49)，得

$$P\left(\inf_{\gamma\in A^c}D(\gamma,L_2)\leqslant 0\right)<\varepsilon/2. \tag{2.50}$$

最后，由(2.16)、(2.17)、(2.38)和(2.50)，得

$$P(\|\hat{\beta}_{n0}\|\geqslant L)<\varepsilon,\ \text{当}\ L\geqslant \max(L_1,L_2).$$

这就证明了(2.14)(注意：已设 $\beta_0=0$，因而 $\beta_{n0}=0$)．如前所述，这也就证明了当 $n\to\infty$ 时，$\hat{\beta}_n$ 依概率收敛于 $\beta_0=0$. 定理2.1证毕. ▌

§2.3 弱相合的必要条件

如果说定理2.1对 ρ 为凸时的 M 估计 $\hat{\beta}_n$ 的弱相合的充分条件作了基本的解决(当然，这方面也还可以提出进一步的问题，例如，若把定理2.1的条件1°～3°加强或削弱一些，则条件(2.3)也应相应地有所变动．这一点在下面论及 LSE 时，可清楚地看到)，则对 $\hat{\beta}_n$ 相合的必要条件问题，除了 LSE 这个特例以外，至今所得的结果很少．本节将对现状作一简略的介绍．由于问题尚在发展过程中，目前所得的结果(除 LSE 外)还有待改进，而证明又颇繁复，所以我们在此将略去大部分证明的细节.

一、最小二乘估计(LSE)

LSE 是 M 估计最重要的特例，其相合性条件问题的研究也最早．最初考虑的情况是模型(1.1)中的随机误差$\{e_i\}$满足 $Ee_i=0$ ($i\geqslant1$)及 Gauss-Markov(缩写为 GM)条件.

$$Ee_ie_j=\sigma^2\delta_{ij},\ 0<\sigma^2<\infty,\ \delta_{ij}=\begin{cases}1,\text{当}\ i=j;\\0,\text{当}\ i\neq j.\end{cases}$$

在这种条件下，很容易证明 β_0的 LSE $\hat{\beta}_n$ 均方相合(因而为弱相合)的充分条件是(2.3)．反面的问题是：在 $Ee_i=0$及 GM 条件下，

(2.3)是否为 LSE 相合的必要条件的问题.这个问题首先由 Drygas 在1976年解决了(见文献[27]),他证明了(2.3)确是必要的.本书作者之一也独立地给出了另一个证明(见文献[21]).事实上有更一般的结果:对任何常向量 c,$c'\beta_0$的 LSE $c'\hat{\beta}_n$ 相合的充要条件是(2.6).

若假定 e_i 为 iid.,则可得到如下更一般的结果:

定理2.6 设在模型(1.1)中,误差 $e_1,e_2,\cdots$为 iid.,且非退化,则当 $c'\beta_0$的 LSE $c'\hat{\beta}_n$ 为相合时,必有(2.6).

证 按 LSE 的周知的公式,有

$$\begin{aligned} c'\hat{\beta}_n &= c'S_n^{-1}\sum_{i=1}^n x_i Y_i \\ &= c'\beta_0 + c'S_n^{-1}\sum_{i=1}^n x_i e_i \\ &\equiv c'\beta_0 + \sum_{i=1}^n c_{ni} e_i, \end{aligned}$$

式中 $$c_{ni} = c'S_n^{-1}x_i.$$

于是由 $c'\hat{\beta}_n$ 为 $c'\beta_0$的相合估计,知

$$\sum_{i=1}^n c_{ni}e_i \xrightarrow{P} 0.$$

由对$\{e_i\}$的假定,根据文献[1]的引理2,应有

$$\sum_{i=1}^n c_{ni}^2 \to 0. \tag{2.51}$$

但由 $c_{ni}=c'S_n^{-1}x_i$ 知

$$\sum_{i=1}^n c_{ni}^2 = c'S_n^{-1}c.$$

由此及(2.51),即得(2.6),定理证毕. ▌

这个结果的特异之处在于并未假定 $Ee_i=0$,甚至连 e_i 的期望也不需存在,更不用说方差了.这个结果启发我们提出如下的问题:为着 LSE $\hat{\beta}_n$ 为相合,是否必须要求 e_i 的期望或方差存在以及 e_i 的期望是否必须为0呢?对这问题的回答如下:

（1）就特例而言，存在着这样的情况，其中 e_i 的方差甚至期望不存在，但 $\hat{\beta}_n$ 为相合. 例如取 $p=1, x_1=x_2=\cdots=1$（β_0为位置参数），则根据大数定律，只要 e_i 为 iid. 且 $Ee_1=0$，则 β_0的 LSE 即 $\hat{\beta}_n=(Y_1+\cdots+Y_n)/n$ 为 β_0的相合估计，而无须 e_1的方差有限. 又如仍取 $p=1, x_i=(-1)^{i-1}$ $(i\geqslant 1)$，e_i 为 iid.，e_1的期望存在，但不必为0，则 $\hat{\beta}_n$ 仍为相合的. 最后，取 $p=1$，e_i 为 iid.，e_1的分布为 Cauchy 分布，e_1的期望也不存在. 取 $x_i=i^2$，则 β_0的 LSE 为

$$\hat{\beta}_n=\Big(\sum_{i=1}^n i^4\Big)^{-1}\sum_{i=1}^n i^2Y_i=\beta_0+\Big(\sum_{i=1}^n i^4\Big)^{-1}\sum_{i=1}^n i^2e_i,$$

因为 e_i 为 Cauchy 分布，有 $M\equiv E|e_i|^{1/2}<\infty$，故

$$E\Big|\Big(\sum_{i=1}^n i^4\Big)^{-1}\sum_{i=1}^n i^2e_i\Big|^{1/2}\leqslant\Big(\sum_{i=1}^n i^4\Big)^{-1/2}\sum_{i=1}^n E(i^2|e_i|)^{1/2}$$

$$=\Big(\sum_{i=1}^n i^4\Big)^{-\frac{1}{2}}\sum_{i=1}^n iM\to 0\quad(\text{当 } n\to\infty).$$

这证明了$\Big(\sum_{i=1}^n i^4\Big)^{-1}\sum_{i=1}^n i^2e_i\xrightarrow{P}0$，因而 $\hat{\beta}_n$ 为相合.

（2）如果要维持(2.3)的充分性，则在 e_i 为 iid. 的前提之下，$Ee_i=0$和 $Ee_i^2<\infty$这两个条件一个也不可缺少. 即可以证明以下定理：

定理2.7　任给自然数 $p\geqslant 1$及一列 iid. 变量 $e_1,e_2,\cdots$，而 $Ee_i=0$和 $Ee_i^2<\infty$这两个条件至少有一个不成立，则可以找到一列 p 维向量$\{x_i\}$满足(2.3)，但在线性模型 $Y_i=x_i'\beta_0+e_i$ $(1\leqslant i\leqslant n)$之下，$\beta_0$的 LSE 不是相合的.

证明不在此给出了. 在这个意义上可以说，为了使(2.3)保持为充分条件，$Ee_i=0$和 $Ee_i^2<\infty$这两个条件都不可减弱.

（3）前面的特例已表明：如果我们把加在$\{x_i\}$上的条件收紧些，则为 $\hat{\beta}_n$ 相合，从而加在 e_i 上的条件可以适当减弱. 对一个重要情况已有了完满的结果如下：

考虑线性模型(1.1)，设误差 $e_1,e_2,\cdots$为 iid.，满足条件

$$Ee_1=0,\ 0<E|e_1|^r<\infty\quad(\text{对某个 } r\in[1,2)).\tag{2.52}$$

对任一 p 维常向量 c，$c'\beta_0$的 LSE $c'\hat{\beta}_n$ 有形状

$$c'\hat{\beta}_n = u_{n1}Y_1 + \cdots + u_{nn}Y_n,$$

其中 $u_{ni}=c'S_n^{-1}x_i$，关于 $c'\hat{\beta}_n$ 的相合性条件已彻底解决了，把 $|u_{n1}|$，$\cdots$，$|u_{nn}|$ 按由大到小排列为 $u_{n(1)}\geqslant\cdots\geqslant u_{n(n)}$.

定义

$$W_r(n)=\max_{1\leqslant i\leqslant n} i u_{n(i)}^r,$$

$$V(n) = \max_{1\leqslant j\leqslant n}\left|\sum_{i=1}^{n} u_{ni} I(|u_{ni}| \geqslant u_{n(j)})\right|.$$

定理2.8 在条件(2.52)之下，$c'\hat{\beta}_n$ 相合的充要条件为：对 $1<r<2$，以下两条件同时成立：

$$c'S_n^{-1}c \to 0, \tag{2.53}$$

$$W_r(n) = O(1). \tag{2.54}$$

对 $r=1$，则除以上两条件以外，还须加上

$$V(n) = O(1). \tag{2.55}$$

这个定理的证明很繁，限于篇幅，此处不给出了. 应当指出的是，必要性是在下述意义下而言的：若所提的条件不全成立，则必存在一维分布 F_0，满足 $\int_{-\infty}^{\infty} x\mathrm{d}F_0 = 0$, $0 < \int_{-\infty}^{\infty} |x|^r\mathrm{d}F_0 < \infty$，使如在模型(1.1)中误差 $e_1,e_2,\cdots$为 iid.，且有公共分布 F_0时，$c'\hat{\beta}_n$ 不依概率收敛于 $c'\beta_0$. 必要性的这种解释，在本章引言部分的末尾已提到过了. 如果在(2.52)中 $r=2$，则单是(2.53)这个条件就构成充要条件了[且必要部分还有更确定的意义：只要(2.53)不成立，则不管 e_i 的公共分布如何(但满足(2.52))，$c'\hat{\beta}_n$ 必不依概率收敛于 $c'\beta_0$，而不只是对个别例外的 F_0而言. 可以证明，对 $r<2$，具有这样明确意义的充要条件则不存在]. 因此，对(2.54)和(2.55)可以理解为是由于对 e_i 的假定的放松(由要求 $Ee_i^2<\infty$改为要求 $E|e_i|^r<\infty$)而在$\{x_i\}$的条件上所付出的代价.

回到基本定理2.1，还可以作如下的解释：把对 e_i 的要求由 $E|e_i|^2<\infty$减弱为$E|e_i|^r<\infty$，相当于定理2.1的条件3°减弱为：

$$E|\Psi(e_1 \pm \Delta)|^r < \infty \quad (\text{对于某个 } r \in [1,2)). \tag{2.56}$$

我们看到了：至少在 LSE 这一特例上，(2.3)已不再是 M 估计相

合的充分条件了. 在条件(2.56)之下,M 估计相合的充分条件问题现在尚未解决.

二、最小绝对偏差估计(LADE)

LADE 是除 LSE 之外的 M 估计的另一个重要的特例,与 LSE 一样,对 LADE 相合性的必要条件的研究注意力主要也是集中在$\{x_i\}$上,即在对$\{e_i\}$作一定的假定下去寻求为使 LADE $\hat{\beta}_n$ 为相合时$\{x_i\}$所必须满足的条件. 自然,首要的兴趣乃是集中在条件(2.3)上.

首先要仔细分析一下,关于$\{e_i\}$的假定要怎样提才合理?在定理2.1的三个条件中,对 LADE 而言,条件3°自然满足,因为$|\Psi|\leqslant 1$,条件1°归于要求 $\text{med}(e_1)=0$,这个条件可以留下来,因若 $\text{med}(e_1)\neq 0$,则如以前所分析过的,用 LADE 不合适,问题就在于条件2°. 不难验证,对 LADE 这个特例,条件2°有形式如下:

定义
$$C(x)=\begin{cases}P(0<e_1<x)/x, x>0;\\ P(x<e_1<0)/|x|, x<0,\end{cases}$$
则有
$$P(e_1\leqslant 0)=1/2\Rightarrow\liminf_{x\downarrow 0} C(x)>0;$$
$$P(e_1\geqslant 0)=1/2\Rightarrow\liminf_{x\uparrow 0} C(x)>0.$$
粗略地说, e_1的概率在0的邻域内应有一个不低于线性程度的聚集,聚集度愈高,则 LADE 愈倾向于相合,因如考虑一个极端的情况,即 e_1退化到0,则只要 S_n 满秩,LADE $\hat{\beta}_n$ 就以概率1取真参数β_0为值,故当然为相合. 因此,要获得合理的必要条件,尤其是如要使(2.3)可能成为必要条件,则要限制 e_1的概率在0的邻域内的聚集,具体地说,不超过线性程度. 定义
$$\tilde{c}(x)=\begin{cases}P(0\leqslant e_1<x)/x,\ x>0;\\ P(x<e_1\leqslant 0)/|x|,\ x<0.\end{cases}$$
则
$$\limsup_{x\to 0}\tilde{c}(x)<\infty. \tag{2.57}$$

总结来说:LADE $\hat{\beta}_n$ 相合的必要条件的提法就是在假定 e_i 为 iid.,$\text{med}(e_1)=0$,且(2.57)成立的前提下,企图证明(2.3)是 $\hat{\beta}_n$

相合的必要条件.

这个问题离彻底解决还差很远，目前只有一些初步的结果，现在不加证明地引述如下：

1° 对 $p=1$的情况已得到证明；

2° 对一般的 p，在一些补充条件下得到证明，条件是：$\{x_i\}$有界，e_1在0的邻域$(-\delta,\delta)$内有概率密度 f，f 在$(-\delta,\delta)$内绝对连续，且 $\int_{-\delta}^{\delta}(f(x))^{-1}(f'(x))^2\mathrm{d}x<\infty$. 以上两个结果的证明参看文献[25].

3° 对一般的 p，证明了一个较弱的必要条件：

$$\sum_{i=1}^{\infty}\|x_i\|^2=\infty. \tag{2.58}$$

如以 $\bar{\lambda}_n$ 和 $\underline{\lambda}_n$ 分别记 S_n 的最大和最小特征根，则(2.3)和(2.58)分别相当于 $\underline{\lambda}_n\to\infty$和 $\bar{\lambda}_n\to\infty$，由此可看出：条件(2.58)与(2.3)是有一段距离的.

下面我们只给出最后这个结果的证明，首先把推理的主要部分提出来，列成几条引理如下：

引理2.1 设模型(1.1)中 β_0的维数 $p=1$，令 $M_n=\sum_{i=1}^{n}|x_i|$. 定义随机变量 ξ，其分布为：

$$P(\xi=Y_i/x_i)=|x_i|/M_n\quad(1\leqslant i\leqslant n),$$

则 ξ 的任一中位数必是 β_0的 LADE. 反之，β_0的任一个 LADE 必是 ξ 的中位数.

证明很容易.只须注意到由 ξ 的定义，有

$$\sum_{i=1}^{n}|Y_i-x_i\beta|=M_n\sum_{i=1}^{n}|Y_i/x_i-\beta||x_i|/M_n=M_nE|\xi-\beta|,$$

再利用熟知的事实——使 $E|\xi-\beta|$达到最小的 β 值的集与 ξ 的一切中位数的集相同.

为了论述方便，以后用 MED(ξ)记 ξ 的一切中位数的集（而 med(ξ)则是指 ξ 的某一个中位数），这样，用$\{\mathrm{MED}(\xi)<c\}$表示 ξ 的每一中位数都小于 c，等等.

引理2.2　设对固定的 n 有独立随机变量 $\xi_{n1},\cdots,\xi_{nn}$，其分布为

$$P(\xi_{ni}=1)=1-p(\xi_{ni}=0)=p_{ni},$$
$$1/4\leqslant p_{ni}\leqslant 1/2,\ 1\leqslant i\leqslant n.$$

又设 $a_{n1},\cdots,a_{nn}$ 都是实常数. 令

$$\xi_n=\sum_{i=1}^{n}a_{ni}(\xi_{ni}-p_{ni}),$$

则

$$\limsup P(\xi_n\geqslant 0)>0. \tag{2.59}$$

为了证明，不失普遍性，设 $a_{n1},\cdots,a_{nn}$ 不全为0. 定义

$$q_{ni}^2=p_{ni}(1-p_{ni}),\ e_{ni}=(\xi_{ni}-p_{ni})/q_{ni}\quad(1\leqslant i\leqslant n),$$

$$b_{ni}=a_{ni}q_{ni}\Big/\sum_{j=1}^{n}q_{nj}^2a_{nj}^2,\ \eta_n=\sum_{i=1}^{n}b_{ni}e_{ni},$$

则 $e_{n1},\cdots,e_{nn}$ 独立，各有期望0，方差为1. 而(2.59)等价于

$$\limsup P(\eta_n\geqslant 0)>0. \tag{2.60}$$

令 $a=1/14$，$b=15/16$，而

$$\{i_1,\cdots,i_r\}=\{i:1\leqslant i\leqslant n,|b_{ni}|\geqslant a\}.$$

因为 $\sum_{i=1}^{n}b_{ni}^2=1$，有 $r\leqslant 14^2=196$，分两种情况讨论：

(1) $B_n\equiv b_{ni_1}^2+\cdots+b_{ni_r}^2\geqslant b$

这时，注意到 $b_{n1}^2+\cdots+b_{nn}^2=1$，因而 $|b_{ni}|\leqslant 1$，有 $|b_{ni_1}|+\cdots+|b_{ni_r}|\geqslant b$. 令

$$\theta_{n1}=b_{ni_1}e_{ni_1}+\cdots+b_{ni_r}e_{ni_r},\theta_{n2}=\eta_n-\theta_{n1}.$$

因为 $p_{ni}\leqslant 1/2$，有 $(1-p_{ni})/q_{ni}\geqslant 1$，故

$$P(\theta_{n1}\geqslant b)\geqslant P(\xi_{ni_1}=1,\cdots,\xi_{ni_r}=1)\geqslant 4^{-r}\geqslant 4^{-196}.$$

又因 $E\theta_{n2}=0$，$\mathrm{Var}(\theta_{n2})=1-B_n\leqslant 1-b$，故

$$P(|\theta_{n2}|\leqslant b/2)\geqslant 1-(1-b)(b/2)^{-2}=161/225.$$

因此

$$P(\eta_n\geqslant 0)\geqslant P(\theta_{n1}\geqslant b,|\theta_{n2}|\leqslant b/2)\geqslant 161/(4^{196}\times 225)>0. \tag{2.61}$$

(2) $B_n<b$

记 $\theta_{n3}=\theta_{n2}/\sqrt{1-B_n}$，并以 F_n 记其分布函数. 注意：当$1/4\leqslant p_{ni}\leqslant 1/2$时，有

$$
\begin{aligned}
E|e_{ni}|^3 &= (1-p_{ni})p_{ni}^3q_{ni}^{-3}+p_{ni}(1-p_{ni})^3q_{ni}^{-3}\\
&=(p_{ni}^2+(1-p_{ni})^2)/q_{ni}<2,
\end{aligned}
$$

又若记 $T=\{1,\cdots,n\}-\{i_1,\cdots,i_r\}$，则 $\sum_{i\in T}|b_{ni}|^3\leqslant a(1-B_n)$. 于是，注意$1-B_n>1-b=1/16$，利用 Berry-Esseen 不等式(见本书第6章中引理6.1)，有

$$
\begin{aligned}
\sup_x|F_n(x)-\Phi(x)| &\leqslant 0.8\times(1-B_n)^{-3/2}\cdot 2a(1-B_n)\\
&<0.8\times 4\times 2/14=6.4/14,
\end{aligned}
$$

其中 $\Phi(x)$为 $N(0,1)$的分布函数. 由此可知

$$
\begin{aligned}
P(\theta_{n2}>0)=P(\theta_{n3}>0) &= 1-F_n(0)\\
&\geqslant 1-\Phi(0)-6.4/14=3/70.
\end{aligned}
$$

又 $P(\theta_{n1}\geqslant 0)\geqslant 4^{-196}$，于是有

$$
P(\eta_n>0)\geqslant P(\theta_{n1}\geqslant 0,\ \theta_{n2}>0)\geqslant(3/70)4^{-196}>0,
\tag{2.62}
$$

结合(2.61)和(2.62)，即得(2.60). 从而证明了本引理. ▌

引理2.3 设 $e_1,e_2,\cdots$ 独立，各有中位数0，且满足条件(2.57). 又设 g_{ni}、h_{ni} $(1\leqslant i\leqslant n,n\geqslant 1)$都是非负常数，且

$$
G_n\equiv\sum_{i=1}^n g_{ni}>0,\ H_n^2\equiv\sum_{i=1}^n h_{ni}^2\leqslant H<\infty\quad(n\geqslant 1).
\tag{2.63}
$$

给定常数列$\{\varepsilon_n\}$，$\varepsilon_n\downarrow 0$，定义

$$
p_n=\sum\nolimits_Q P(e_i>h_{ni}\varepsilon_n(\text{当 } i\in Q),\ e_i\leqslant h_{ni}\varepsilon_n(\text{当 } i\in Q^c)),
\tag{2.64}
$$

此处$\sum_Q$ 表示求和遍及一切满足下述条件的集 Q：

$$
Q\subset\{1,\cdots,n\},\ \sum\nolimits_{i\in Q}g_{ni}>G_n/2,\ (Q^c=\{1,\cdots,n\}-Q),
$$

则有

$$\limsup p_n > 0. \tag{2.65}$$

为了证明，不妨设 $G_n=1$. 令

$$\xi_{ni} = I(e_i > h_{ni}\varepsilon_n) \quad (1 \leqslant i \leqslant n),\ \xi_n = \sum_{i=1}^n g_{ni}\xi_{ni}.$$

易见：为证(2.65)，只须证

$$\limsup P(\xi_n > 1/2) > 0. \tag{2.66}$$

记

$$p_{ni} = E\xi_{ni} = P(e_i > h_{ni}\varepsilon_n) \quad (1 \leqslant i \leqslant n).$$

则由本引理的假设、(2.63)式及 $h_{ni}\geqslant 0, \varepsilon_n \downarrow 0$，知当 n 充分大时，有 $1/4\leqslant p_{ni}\leqslant 1/2 \ (1\leqslant i\leqslant n)$. 又

$$\begin{aligned} 1/2 - E\xi_n &= \sum_{i=1}^n g_{ni}[1/2 - P(e_i > h_{ni}\varepsilon_n)] \\ &\leqslant \sum_{i=1}^n g_{ni}P(0 \leqslant e_i \leqslant h_{ni}\varepsilon_n) \leqslant l_1\sum_{i=1}^n g_{ni}h_{ni}\varepsilon_n \\ &\leqslant l_1K_nH\varepsilon_n, \end{aligned} \tag{2.67}$$

此处 $K_n^2 \equiv \sum_{i=1}^n g_{ni}^2 \leqslant 1$. 因为 $G_n=1$，而 $g_{ni}\geqslant 0$，记

$$D_n^2 = \sum_{i=1}^n g_{ni}^2 p_{ni}(1 - p_{ni}),$$

则有

$$1 \geqslant D_n^2 \geqslant K_n^2/16, D_n \geqslant K_n/4. \tag{2.68}$$

以下分两种情况讨论：

(1) $\max\limits_{1\leqslant i\leqslant n} g_{ni}/D_n \to 0$

这时随机变量的三角阵列 $\{g_{ni}(\xi_{ni}-p_{ni}): 1\leqslant i\leqslant n, n\geqslant 1\}$ 满足所谓uan(一致渐近可忽略)条件. 根据中心极限定理，有

$$(\xi_n - E\xi_n)/D_n \xrightarrow{\mathscr{L}} N(0,1) \quad (n \to \infty).$$

由此及(2.67)和(2.68)，有

$$\begin{aligned} P(\xi_n > 1/2) &= P[(\xi_n - E\xi_n)/D_n > (1/2 - E\xi_n)/D_n] \\ &\geqslant P[(\xi_n - E\xi_n)/D_n > 4l_1\varepsilon_n H] \to 1/2. \end{aligned}$$

这证明了(2.66).

(2) $\max\limits_{1\leqslant i\leqslant n} g_{ni}/D_n \not\rightarrow 0$

必要时取子列并调整足标编号，不失普遍性，设存在 $\varepsilon>0$，使当 n 充分大时，有

$$g_{n1}=\max_{1\leqslant i\leqslant n} g_{ni}\geqslant \varepsilon D_n.$$

记 $\theta_n=\sum\limits_{i=2}^{n} g_{ni}(\xi_{ni}-p_{ni})$. 因为 $\varepsilon_n\rightarrow 0$，当 n 充分大时，有 $4l_1\varepsilon_n H<\varepsilon/4$. 故当 n 充分大时，有

$$\begin{aligned}
P(\xi_n>1/2)&=P(\xi_n-E\xi_n>1/2-E\xi_n)\\
&\geqslant P(\xi_n-E\xi_n>l_1K_nH\varepsilon_n)\geqslant P(\xi_n-E\xi_n>\varepsilon D_n/4)\\
&\geqslant P(g_{n1}(\xi_{n1}-p_{n1})>\varepsilon D_n/4)P(\theta_n\geqslant 0)\\
&\geqslant P(\xi_{n1}-p_{n1}>1/4)P(\theta_n\geqslant 0)\\
&\geqslant P(\xi_{n1}=1)P(\theta_n\geqslant 0)\geqslant 4^{-1}P(\theta_n\geqslant 0).
\end{aligned}$$

按引理2.2，有 $\limsup P(\theta_n\geqslant 0)>0$，故由上式仍得出(2.66). 引理证毕. ▌

现在转到3°的证明：不妨设 $\beta_0=0$，故 $Y_i=e_i$. 记 $\hat{\beta}_n$ 的 j 分量为 $\hat{\beta}_{nj}$，x_i 的 j 分量为 x_{ij}，不妨设 $x_1,\cdots,x_n$ 不全为0，且不失普遍性，设 $x_{1p},\cdots,x_{np}$不全为0，且皆为非负. 事实上，若某个 $x_{ip}<0$. 则把相应的方程 $Y_i=x_i'\beta_0+e_i$ 改成 $-Y_i=(-x_i)'\beta_0+(-e_i)$，容易看到，若 $\{e_i\}$满足条件(2.57)，则不论+、-号如何取，$(\pm e_1,\pm e_2,\cdots)$必也满足(2.57). 这样 x_i 可换为 $-x_i$，其相应的 p 分量 $-x_{ip}>0$.

记 $A_n=x_{1p}+\cdots+x_{np}$，定义随机变量 η_n，分布为

$$P(\eta_n=(e_i-x_i'\hat{\beta}_n)/x_{ip})=x_{ip}/A_n\quad(1\leqslant i\leqslant n),$$

注意：在此处的 η 及下面的 η^* 和 θ_n 的定义中，将 e_i 和 $\hat{\beta}_n$ 看作为常数. 往下证明

$$\operatorname{med}(\eta_n)=0.\tag{2.69}$$

为此令

$$Z_i=e_i-\sum_{j=1}^{p-1}x_{ij}\hat{\beta}_{nj}\quad(1\leqslant i\leqslant n),$$

定义 $H(t)=\sum\limits_{i=1}^{n}|Z_i-x_{ip}t|$，由 $\hat{\beta}_n$ 为 β_0的 LADE 知，$\hat{\beta}_{np}$应为 $H(t)$

的最小值点，故若定义随机变量 η_n^*，其分布为

$$P(\eta_n^* = Z_i/x_{ip}) = x_{ip}/A_n \quad (1 \leqslant i \leqslant n),$$

则按引理2.1，应有 $\hat{\beta}_{np}=\text{med}(\eta_n^*)$. 但 η_n 与 $\eta_n^*-\beta_{np}$同分布，因而证明了(2.69).

现设$\{\epsilon_n\}$为常数列，$\epsilon_n \downarrow 0$，记 $L_n^2=\sum_{i=1}^{n}\|x_i\|^2$. 定义随机变量 θ_n，分布为：

$$P(\theta_n = (e_i - \|x_i\|\epsilon_n/L_n)/x_{ip}) = x_{ip}/A_n \quad (1 \leqslant i \leqslant n).$$

如果 $L_n\|\hat{\beta}_n\|\leqslant\epsilon_n$，则有

$$\begin{aligned} e_i - x_i'\hat{\beta}_n &\geqslant e_i - \|x_i\|\|\hat{\beta}_n\| \\ &\geqslant e_i - \epsilon_n\|x_i\|/L_n \quad (1 \leqslant i \leqslant n), \end{aligned}$$

故由(2.69)知 $\text{med}(\theta_n)\leqslant 0$(此式的意义是：$\theta_n$ 有一个中位数为0). 这证明了

$$\{L_n\|\hat{\beta}_n\| \leqslant \epsilon_n\} \subset \{\text{med}(\theta_n) \leqslant 0\}.$$

故若以 $\text{MED}(\theta_n)$记 θ_n 的一切中位数的集，则有

$$\{L_n\|\hat{\beta}_n\| > \epsilon_n\} \supset \{\text{MED}(\theta_n) \leqslant 0\}^c = \{\text{MED}(\theta_n) > 0\}. \tag{2.70}$$

记 $h_{ni}=\|x_i\|/L_n, g_{ni}=x_{ip}/A_n \quad (1\leqslant i\leqslant n)$，则满足了(2.63). 并且 $G_n=H_n^2=1$. 因为按 θ_n 的定义，易见 $P(\text{MED}(\theta_n)>0)\geqslant p_n$，$p_n$ 由(2.64)定义，故由(2.70)及引理2.3，得

$$\lim\sup P(L_n\|\hat{\beta}_n\| > \epsilon_n) > 0.$$

因为上式对任何 $\epsilon_n \downarrow 0$都成立，故得到

$$L_n\hat{\beta}_n \neq O_p(1). \tag{2.71}$$

由此可知，若 $\sum_{i=1}^{\infty}\|x_i\|^2<\infty$，则 $\hat{\beta}_n\neq O_p(1)$，即 $\hat{\beta}_n$ 不依概率收敛于 $0=\beta_0$. 从而证明了所要的结果.

所证明的上述结果与猜想目标($S_n^{-1}\to 0$为必要)之间仍存在距离，这是因为由 $S_n^{-1}\to 0$，可推出 $\sum_{i=1}^{\infty}\|x_i\|^2=\infty$，但其逆不真.

这个结果可略加推广为下述形式：考虑模型

$$Y_{ni} = x_{ni}'\beta_{n0} + e_i \quad (1 \leqslant i \leqslant n, n \geqslant 1),$$

其中 $x_{n1},\cdots,x_{nn}$ 为常向量，随机误差满足前述条件. 以 $\hat{\beta}_n$ 记 β_{n0} 的 LADE，则有

$$\sum_{i=1}^{n} \|x_{ni}\|^2 (\hat{\beta}_n - \beta_{n0}) \neq O_p(1). \qquad (2.72)$$

特别是，取 $x_{ni} = S_n^{-1/2} x_i$，$\beta_{n0} = S_n^{1/2}\beta_0$，则 β_{n0} 的 LADE 为 $\hat{\beta}_n = S_n^{1/2}\hat{\beta}_n$，其中 $\hat{\beta}_n$ 为模型(1.1)中 β_0 的 LADE. 注意到 $\sum_{i=1}^{n} \|x_{ni}\|^2$ 等于 β_0 的维数 p，由(2.72)得

$$S_n^{1/2}(\hat{\beta}_n - \beta_0) \neq O_p(1). \qquad (2.73)$$

另一方面，在第2章中曾经证明(见(2.14)式)了在一定条件下有 $S_n^{1/2}(\hat{\beta}_n - \beta_0) = O_p(1)$. 把这与(2.73)结合起来，就得出：在两方面的条件都满足时，$S_n^{1/2}(\hat{\beta}_n - \beta_0)$ 的确切的阶就是 $O_p(1)$，即它能达到这个阶，但不能更高. 例如，在 $e_1, e_2, \cdots$ 为 iid.，其公共分布在0点的一邻域内有密度 f，且 f 在此邻域内有正的下界及有限的上界时，就是这样一种情况.

另一方面，若(2.57)不成立，则可以证明：(2.3)不是 $\hat{\beta}_n$ 相合的必要条件. 这一点是不难理解的：(2.57)不成立意味着 e_i 的概率分布高度集中在0点附近. 这一情况有利于 β_0 的估计，即增加其估计(例如 LADE $\hat{\beta}_n$)的相合的可能性，因而对 $\{x_i\}$ 的要求可相应降低.

为证明上述断言，定义

$$c_1(x) = 1/2 - P(e_1 \geqslant x) \quad (\text{当 } x > 0),$$
$$c_1(x) = 1/2 - P(e_1 \leqslant x) \quad (\text{当 } x \leqslant 0).$$

因为(2.57)不成立，故有

$$\limsup_{|x| \downarrow 0} c_1(x)/|x| = \infty. \qquad (2.74)$$

事实上，若 $P(e_1 = 0) = 0$，则 $c_1(x) = c(x)$，当(2.57)不成立时，当然有(2.74). 若 $P(e_1 = 0) > 0$，则或者有 $P(e_1 > 0) < 1/2$，这时 $\limsup_{x \downarrow 0} c_1(x)/x = \infty$；或者 $P(e_1 < 0) < 1/2$，这时 $\limsup_{x \uparrow 0} c_1(x)/|x| = \infty$. 为确定计，设

$$\limsup_{x \downarrow 0} c_1(x)/x = \infty,$$

则易见可找到一串正数$\{t_i\}$,使

$$\sum_{i=1}^{\infty} t_i c_1(t_i) = \infty, \quad \sum_{i=1}^{\infty} t_i^2 < \infty.$$

由此可找到一串正数 $d_i \to \infty$,使

$$\sum_{i=1}^{\infty} d_i^2 t_i^2 < \infty.$$

记 $\tilde{t}_i = d_i t_i$, $\tilde{\varepsilon}_i = 1/d_i$,则有

$$\infty = \sum_{i=1}^{\infty} t_i c_1(t_i) \Rightarrow \sum_{i=1}^{\infty} \tilde{t}_i c_1(\tilde{\varepsilon}_i \tilde{t}_i) = \infty.$$

现在令 $x_{2i-1} = -\tilde{t}_i$, $x_{2i} = \tilde{t}_i (i \geqslant 1)$,而考虑线性模型

$$Y_i = x_i \beta_0 + e_i \quad (1 \leqslant i \leqslant n),$$

记 $\varepsilon_{2i-1} = \varepsilon_{2i} = \tilde{\varepsilon}_i \quad (i \geqslant 1)$,则有 $\varepsilon_i \to 0$ (当 $i \to \infty$),且

$$\sum_{i=1}^{\infty} x_{2i} c_1(\varepsilon_{2i} x_{2i}) = \sum_{i=1}^{\infty} |x_{2i-1}| c_1(\varepsilon_{2i-1} |x_{2i-1}|) = \infty,$$

$$\sum_{i=1}^{\infty} x_i^2 < \infty. \tag{2.75}$$

不失普遍性,设 $\beta_0 = 0$,以 $\hat{\beta}_n$ 记 β_0的 LADE. 往下证明 $\hat{\beta}_n \xrightarrow{P} 0$,因而 $\hat{\beta}_n$ 为相合.

任给 $\varepsilon > 0$,记

$$Z_n = \sum_{i=1}^{n} M_n^{-1} |x_i| I(e_i/x_i \geqslant \varepsilon),$$

$$\tilde{Z}_n = \sum_{i=1}^{n} M_n^{-1} |x_i| I(e_i/x_i \leqslant -\varepsilon),$$

其中 $M_n = \sum_{i=1}^{n} |x_i|$. 则按引理2.1,有

$$\{\hat{\beta}_n \geqslant \varepsilon\} \subset \{\mathrm{MED}(\xi) < \varepsilon\}^c \subset \{Z_n \geqslant 1/2\},$$

$$\{\hat{\beta}_n \leqslant -\varepsilon\} \subset \{\mathrm{MED}(\xi) > -\varepsilon\}^c \subset \{\tilde{Z}_n \geqslant 1/2\},$$

此处的 ξ 就是引理2.1中所定义的随机变量. 故为证 $\hat{\beta}_n \xrightarrow{P} 0$,只

需证

$$P(Z_n \geqslant 1/2) \to 0, \quad P(\tilde{Z}_n \geqslant 1/2) \to 0.$$

以前一式子为例，按 $c_1(x)$ 的定义及 $\mathrm{MED}(e_1)=0$，有 $c_1(x) \geqslant 0$，且当 $0<u<v$ 时，有 $c_1(u) \leqslant c_1(v)$. 由这些事实以及 $\varepsilon>0$ 固定同时 $\varepsilon_i>0, \varepsilon_i \to 0$，有

$$M_n(1/2 - EZ_n) = \sum_{i=1}^{n} |x_i| c_1(\varepsilon x_i) \geqslant \sum_{i=1}^{[n/2]} x_{2i} c_1(\varepsilon x_{2i})$$

$$\geqslant \sum_{i=1}^{[n/2]} x_{2i} c_1(\varepsilon_{2i} x_{2i}) \to \infty$$

（当 $n \to \infty$）（见(2.75)）.

又根据(2.75)，有

$$\mathrm{Var}(M_n Z_n) \leqslant \sum_{i=1}^{n} x_i^2 \leqslant \sum_{i=1}^{\infty} x_i^2 < \infty.$$

由此可知，当 $n \to \infty$ 时，有

$$P(Z_n \geqslant 1/2) = P(M_n(Z_n - EZ_n) \geqslant M_n(1/2 - EZ_n))$$

$$\leqslant \mathrm{Var}(M_n Z_n)/(M_n(1/2 - EZ_n))^2 \to 0.$$

类似地可证明

$$P(\tilde{Z}_n \geqslant 1/2) \to 0.$$

如前所述，即证明了 β_0 的 LADE 为相合的. 但 $S_n^{-1} = \left(\sum_{i=1}^{n} x_i^2\right)^{-1}$，根据(2.75)，此值当 $n \to \infty$ 时，并不收敛于0. 这说明对这个 $\{e_i\}$ 而言（它使(2.57)不满足），(2.3)并非 LADE 相合的必要条件.

三、ρ 为一般凸函数的情况

考虑到定理2.1，并参照本节第二段中对 LADE 情况所作的分析，对 ρ 为一般凸函数情况下的 M 估计 $\hat{\beta}_n$ 相合性的必要条件，最理想的结果应有如下的形式：

命题 若在模型(1.1)中，随机误差 $e_1, e_2, \cdots$ 为 iid.，且存在函数 Ψ （$\Psi_- \leqslant \Psi \leqslant \Psi_+$）及常数 $c>0, \Delta>0$，使

$$|E\Psi(e_1 + u)| \leqslant c|u|, \ |u| < \Delta, \tag{2.76}$$

则当 $\hat{\beta}_n$ 为 β_0 的弱相合估计时，必有(2.3).

这个命题中的要点是：条件(2.76)把条件(2.1)反过来了，其所以要反过来，理由与在LADE的情况下引出条件(2.57)的道理相同. 注意由于条件(2.76)，必有 $E\Psi(e_1)=0$，因此这条件不需要另行提出. 至于定理2.1的条件3°完全删去了(注意在LADE的情况，这条件是自动满足的)，是基于在LSE情况下的结果的考虑，即：在定理2.5中并不需要假定误差方差为有限的.

到现在为止，这个命题还只是一个猜测. 虽然有许多理由相信它是正确的，其严格的证明可能极为困难. 目前，仅仅对 $p=1$ 的特例并且作了若干附加的假定，证明了(2.3)是 $\hat{\beta}_n$ 弱相合的必要条件.

第 3 章

M 估计的强相合性

沿用前章的记号，考虑线性模型(1.1)，选定函数 ρ，按(1.5)式定义 β_0 的 M 估计 $\hat{\beta}_n$. 本章的目的是探讨使

$$\hat{\beta}_n \to \beta_0, \text{a.s.}$$

成立的条件. 我们只讨论随机误差 $e_1, e_2, \cdots$ 独立的情况，并为了免去一些形式上的麻烦，而进一步假定 $e_1, e_2, \cdots$ 为 iid.（虽然有些结果不难推广到 $e_1, e_2, \cdots$ 独立但不必为同分布的情形）.

同讨论弱相合问题一样，我们的架构是在对函数 ρ 和 e_i 的公共分布作一定的假定之下，去探讨当 $\{x_i\}$ 满足何种条件时，$\hat{\beta}_n$ 有强相合性. 一个自然的选择是定理 2.1 的假定 1°～3°，关于这些条件的意义我们在上一章中已有充分的分析了.

对 M 估计的一个重要特例——LSE，定理 2.1 的条件 1°～3° 相当于要求 $Ee_1=0, Ee_1^2<\infty$. 在这样的假定（实际上在比这更宽一些的假定）下，黎子良等于 1979 年在文献[44]中证明了一个基本结果：(2.3)是 $\hat{\beta}_n$ 强收敛的充分条件. 根据 Drygas 的结果(2.3)也是必要的，因而是充分必要的. 这个结果使我们产生一种奢望：对一般的凸函数 ρ，在定理2.1的假定 1°～3°之下，莫非(2.3)仍是 $\hat{\beta}_n$ 强相合的充分条件？倘若情况果然如此，则(2.3)极有可能也是必要条件（把(2.1)式中的不等号要反过来，参见 § 2.3 第三段中的解释），因而 $\hat{\beta}_n$ 的强相合问题也就有了一个完满的解决. 可惜的是，实际情况却与此相差很远.

首先，迄今为止我们还未能（就一般的凸函数 ρ）在定理2.1的

假定 1°～3°及关于 $\{x_i\}$ 的任何假定下，证得 $\hat{\beta}_n$ 强相合性的结果，需要将假定 3°大大加强，这种加强确是出于必要的，或者是由于证法上的缺陷，迄今还无法推测. 其次，对 LSE 以外的 M 估计，即使像 LADE 这样比较简单的情况，(2.3)也不能保证 $\hat{\beta}_n$ 为强相合. 甚至，不论 S_n^{-1} 以多快的速度趋于 0，也不能保证 $\hat{\beta}_n$ 为强相合.

为证明这一点，任取一串常数$\{A_n\}$，$0<A_n\uparrow\infty$，找一列常数$\{a_i\}$满足以下条件：

$$1<a_1<a_2<\cdots\uparrow\infty;\ S_n\equiv\sum_{i=1}^{n}a_i^2>A_n\quad(n\geqslant 1),\tag{3.1}$$

$$\lim_{n\to\infty}na_n/a_{2n}=0;\ \lim_{n\to\infty}a_n^{-1}\sum_{i=1}^{n-1}a_i=0.\tag{3.2}$$

考察线性模型 $Y_i=x_i\beta_0+e_i\quad(1\leqslant i\leqslant n,n\geqslant 1)$，这里 $x_i=a_i$，e_i 为 iid.，其公共分布有如下的形式：

$$P(e_1=a_k)=P(e_1=-a_k)$$

$$=(k(k+1))^{-1}\quad(k=6,7,8,\cdots),$$

而在区间$(-1/3,1/3)$内，e_1 有概率密度 1，这是一个对称分布，在原点邻域内有非 0 的光滑密度，对 LADE，定理 2.1 的假定 1°～3°当然满足，且因为$|\Psi|\leqslant 1$，Ψ 的任意阶矩甚至其矩母函数也存在，根据(3.1)，$S_n^{-1}<A_n^{-1}$，因此 S_n^{-1} 趋于 0 的速度比给定的 A_n^{-1} 还快. 然而，我们将证明：β_0 的 LADE$\hat{\beta}_n$ 不是强相合的.

不失普遍性，不妨设 $\beta_0=0$，引进事件：

$$Q(N,m)=\{|e_m|\geqslant a_{2N},|e_i|<a_N,1\leqslant i\leqslant m-1\},$$

$$T_r=\bigcup_{m=2^r}^{2^{r+1}}Q(2^{r+1},m).$$

当 $|\beta|\leqslant 1/2$ 时，有事件 $Q(N,m)$发生，且 $m\leqslant N$ 时，有

$$\sum_{i=1}^{m}|Y_i-\beta x_i|=\sum_{i=1}^{m}|e_i-\beta a_i|\geqslant|e_m-\beta a_m|$$

$$\geqslant|e_m|-a_N/2\geqslant|e_m|/2.\tag{3.3}$$

另一方面，

$$\sum_{i=1}^{m}|e_i-\beta a_i|\bigg|_{\beta=e_m/a_m}\leqslant\sum_{i=1}^{m-1}|e_i|+|e_m/a_m|\sum_{i=1}^{m-1}a_i$$

$$\leqslant Na_N+|e_m|a_m^{-1}\sum_{i=1}^{m-1}a_i. \tag{3.4}$$

注意到(3.2)以及当事件 $Q(N,m)$ 发生时，$|e_m|\geqslant a_N$，由上式可知当 N、m 充分大时，(3.4)左边为 $o(|e_m|)$. 故当 N、m 充分大且 $Q(N,m)$ 发生时，由(3.3)和(3.4)可知，β_0 的 LADE $\hat{\beta}_m$ 满足 $|\hat{\beta}_m|>1/2$.

由于对固定的 N，事件 $Q(N,1),\cdots,Q(N,N)$ 互斥，有

$$P(T_r)=\sum_{m=2^r}^{2^{r+1}}P(Q(2^{r+1},m)).$$

根据 e_1 的分布，有

$$P(Q(2^{r+1},m))=(1-2^{-r})^{m-1}2^{-r-1},$$

因此，当 $r\to\infty$ 时，有

$$P(T_r)=\sum_{m=2^r}^{2^{r+1}}2^{-r-1}(1-2^{-r})^{m-1}$$

$$\to e^{-1}(1-e^{-1})/2>0. \tag{3.5}$$

当 T_r 发生时，存在 m $(2^r\leqslant m\leqslant 2^{r+1})$，使 $|\hat{\beta}_m|>1/2$. 由此与(3.5)结合，即知 $\hat{\beta}_n$ 不以概率 1 收敛于 $0=\beta_0$. 从而证明了 $\hat{\beta}_n$ 不是强相合的.

将此例于稍加修改，还可以说明另一个问题：如果把本例中的随机误差 $e_1,e_2,\cdots$ 的公共分布改为 $(-1/2,1/2)$ 内的均匀分布，则

这个改变没有变动原分布的中位数 0，也没有变动原分布在中位数 0 的一个邻域(具体地说即$(-1/3,1/3)$)内的形状，而只是变动了原分布的尾部，但经过这一改变，β_0 的 LADE $\hat{\beta}_n$ 就成为强相合的了. 这个事实很容易利用引理 2.1 来证明. 为此，仍设 $\beta_0=0$，则根据引理 2.1，有 $\hat{\beta}_n=\text{med}(\xi_n)$，其中 ξ_n 有分布：

$$P(\xi_n = e_i/a_i) = a_i/(a_1 + \cdots + a_n) \quad (1 \leqslant i \leqslant n).$$

因为 $|e_i|\leqslant 1/2$ 而 $a_i\to\infty$，知 $\xi_n\to 0$，a. s.. 得所欲证. 这个例子说明了：对于 $\hat{\beta}_n$ 的强相合性而言，不仅随机误差分布在其中位数的邻域内有作用，且整个分布，包括其尾部，也是有影响的，这与 $\hat{\beta}_n$ 的弱相合不同. 对后者而言，只是中位数的一个充分小的邻域内的分布情况才有作用. 从这一点也不难看出，LADE 或更一般的 M 估计，其强相合问题要比弱相合问题复杂得多了.

本章分三节：§ 3.1 和 § 3.3 中分别讨论在函数 ρ 为凸或不必为凸时，M 估计 $\hat{\beta}_n$ 强相合的一些充分条件. 在 § 3.2 中讨论的性质较为特殊，乃是讨论对 LADE 这个特例的一个带有必要条件性质的结果. 这一讨论非常精细，它告诉我们，在 $\hat{\beta}_n$ 的强相合问题中，涉及到必要条件的部分是多么困难.

§ 3.1　ρ 为凸函数的情况

M 估计 $\hat{\beta}_n$ 的相合性取决于两方面的条件：一方面与 ρ 和随机误差的分布有关；一方面则与$\{x_i\}$有关. 当把一方面的条件加强时，另一方面的条件可能减弱. 本节所要讨论的有关 $\hat{\beta}_n$ 强相合的两个结果，即说明了这一点.

仍沿用以前惯常使用的记号，设有模型(1.1)，记 $S_n=x_1x_1'+\cdots+x_nx_n'$，定义

$$d_n = \max_{1\leqslant i\leqslant n} x_i' S_n^{-1} x_i. \tag{3.6}$$

此处假定 S_n^{-1} 存在. 不然的话，估计 β_0 的问题就失掉了意义. 设 ρ 为凸函数，以 Ψ_- 和 Ψ_+ 分别记 ρ 的左、右导数.

定理 3.1　设 $e_1,e_2,\cdots$为 iid.，ρ 为凸函数，满足以下两个条

件：

1° 存在常数 $l_1>0$，$l_2>0$，使

$$E(\rho(e_1+u)-\rho(e_1))\geqslant l_1u^2 \quad (\text{当 } |u|<l_2 \text{ 时}). \tag{3.7}$$

2° 存在常数 $\Delta>0$，使

$$E|\Psi_+(e_1\pm\Delta)|^m\leqslant h_m<\infty \quad (m=1,2,\cdots). \tag{3.8}$$

则当存在 $\delta>0$，使

$$d_n=O(n^{-\delta}) \tag{3.9}$$

时，$\hat{\beta}_n$ 为 β_0 的强相合估计.

在证明此定理之前，我们先来考察一下这几个条件. 不难看出，与定理 2.1 的条件相比，这些条件全面加强了. 首先，按(3.7)，0 是(1.19)式所定义的函数 $g(t)$ 的唯一最小值点. 故按定理 1.4，存在 Ψ，使 $\Psi_-\leqslant\Psi\leqslant\Psi_+$，且有(1.21)，即定理 2.1 的条件 1°成立，且有(1.20). 其次，定理 2.1 的条件 2°也成立. 事实上，为确定计，设 $u>0$，由 ρ 的凸性可知：

$$\rho(e_1+u)-\rho(e_1)=u\Psi(e_1+u).$$

两边取期望，利用(3.7)，即得(2.1)(其中常数 c_0 和 Δ 分别取为 l_1 和 l_2). 对 LADE，易见(3.7)与(2.1)等价. 其次，(3.8)是(2.2)的强化：(2.2)只要求(3.8)当 $m=2$ 时成立，而(3.8)要求对一切 m 成立. 如果说(3.7)强于(2.1)更多是形式的而非本质的，则(3.8)是本质地强于(2.2)：它要求 $\Psi_+(e_1\pm\Delta)$ 有任意阶矩，而不只是二阶矩. 可以猜测，这个条件是不必要地过于强了，但迄今，在改进这一点上尚未有所进展.

现在我们来考察 d_n 这个量，与 S_n 相比，在一定意义上 d_n 这个量更刻划了模型(1.1)的某种本质的东西：设 C 是一个 p 阶非异方阵，我们用线性变换 $\gamma_0=C\beta_0$ 去取代原来的参数向量 β_0，这等于把模型(1.1)变为 $Y_i=z_i'\gamma_0+e_i$(其中 $z_i=C'^{-1}x_i$). 对同一个 ρ，β_0 和 γ_0 的 M 估计 $\hat{\beta}_n$ 和 $\hat{\gamma}_n$ 以关系式 $\hat{\gamma}_n=C\hat{\beta}_n$ 相联系，因此就有关的大样本性质，如相合性、渐近正态性等而言，在变换后的模型或是在原模型中来讨论，都应得出相当的结果. 将变换后的模型的 d_n 值记为 $\tilde{d}_n$，等于

$$\tilde{d}_n = \max_{1\leq i\leq n} x_i'C^{-1}\left(\sum_{i=1}^{n} C'^{-1}x_i x_i' C^{-1}\right)^{-1} C'^{-1} x_i$$
$$= \max_{1\leq i\leq n} x_i'C^{-1}CS_n^{-1}C'C'^{-1}x_i$$
$$= \max_{1\leq i\leq n} x_i'S_n^{-1}x_i = d_n,$$

即线性变换不改变模型的 d_n 值(实际上,不改变每一个 $x_i'S_n^{-1}x_i$ 值). 在以后章节中我们将看到,d_n 这个量在 M 估计大样本性质的讨论中起着重要的作用. 上述讨论有助于我们理解其原因之所在.

条件(2.3)要求 $S_n^{-1}\to 0$. 我们来证明:这个条件弱于 $d_n\to 0$,更不用说要求 d_n 以一定速度趋于 0(如(3.9))了. 为证这一点,先证明一个矩阵不等式:

$$\operatorname{tr}(AB) \geqslant \mu(A)\lambda(B), \tag{3.10}$$

这里 A、B 都是 p 阶非负定方阵,$\mu(A)$、$\lambda(B)$分别是 A 的最大特征根与 B 的最小特征根. 事实上,找正交方阵 P,使

$$PAP' = \operatorname{diag}(a_1,\cdots,a_p), a_1 \geqslant \cdots \geqslant a_p \geqslant 0.$$

则 $a_1=\mu(A)$. 以 b_{ij}记方阵 PBP'的(i,j)元,则易见有 $b_{11}\geqslant\lambda(B)$, $b_{ii}\geqslant 0$. 故有

$$\operatorname{tr}(AB) = \operatorname{tr}(PAP'PBP') = \operatorname{tr}(\operatorname{diag}(a_1,\cdots,a_p)(b_{ij}))$$
$$= \sum_{i=1}^{p} a_i b_{ii} \geqslant a_1 b_{11} \geqslant \mu(A)\lambda(B),$$

即(3.10). 在(3.10)中取 $A=S_n, B=S_m$,使 $n\geqslant m$,而 S_m^{-1} 存在,则 S_m 的最小特征根 $\lambda_m>0$,而 S_n^{-1} 的最大特征根为 λ_n^{-1},于是有:

$$\lambda_m/\lambda_n \leqslant \operatorname{t_r}(S_n^{-1}S_m) = \sum_{i=1}^{m} \operatorname{tr}(S_n^{-1}x_i x_i') = \sum_{i=1}^{m} x_i'S_n^{-1}x_i \leqslant md_n,$$

因此

$$d_n \geqslant \lambda_m/(m\lambda_n) \quad (n \geqslant m). \tag{3.11}$$

若 $d_n\to 0$,则必有 $\lambda_n\to\infty$,即 $S_n^{-1}\to 0$. 反过来,由 $S_n^{-1}\to 0$ 不一定能推出 $d_n\to 0$. 在本章开始处所举的那个例子就属于这种情况. (3.11)也表明了:若 d_n 以某一数量级趋于 0,则 S_n^{-1} 至少以该数量级趋于 0. 另外,d_n 趋于 0 的数量级至多为 $O(1/n)$,这因为

$$nd_n \geqslant \sum_{i=1}^n x_i' S_n^{-1} x_i = \sum_{i=1}^n \mathrm{tr}(S_n^{-1} x_i x_i') = \mathrm{tr}(S_n^{-1} S_n) = \mathrm{tr}(I_p) = p,$$

故 $d_n \geqslant p/n$.

在转到证明定理之前，我们先引进两个有关独立变量和的概率不等式(下列引理 3.1 和引理 3.2)：

引理 3.1(Fuk-Nagaev 不等式) 设随机变量 $\xi_1, \cdots, \xi_n$ 独立，各自有期望 0. 则对任何常数 $t>0$，存在只与 t 有关的常数 $c_{1t}>0$，$c_{2t}>0$，使对任何 $u>0$，有

$$P\left(\left|\sum_{i=1}^n \xi_i\right| \geqslant u\right) \leqslant c_{1t} u^{-t} \sum_{i=1}^n E|\xi_i|^t + 2\exp\left(-c_{2t} u^2 \Big/ \sum_{i=1}^n \mathrm{Var}(\xi_i)\right). \quad (3.12)$$

引理 3.2(Hoeffding 不等式) 设随机变量 $\xi_1, \cdots, \xi_n$ 独立，各自有期望 0，且 $a_i \leqslant \xi_i \leqslant b_i$ $(1 \leqslant i \leqslant n)$，其中 a_i、b_i 都是常数. 则对任何 $u>0$，有

$$P\left(\left|\sum_{i=1}^n \xi_i\right| \geqslant u\right) \leqslant 2\exp\left(-2u^2 \Big/ \sum_{i=1}^n (b_i - a_i)^2\right). \quad (3.13)$$

这两个不等式的证明，分别见参考文献[29]与[31]. 对有界随机变量的情况，Hoeffding 不等式对偏差概率给出一个指数界限，因而很有用. 比 Hoeffding 稍早，Bennett 提出了一个形式稍异但功用基本相同的不等式，也引述如下(引理 3.3)：

引理 3.3(Bennett) 设随机变量 $\xi_1, \cdots, \xi_n$ 独立，各有期望 0，且 $|\xi_i| \leqslant b < \infty$ $(1 \leqslant i \leqslant n$，$b$ 为常数). 记 $\sigma^2 = \sum_{i=1}^n \mathrm{Var}(\xi_i)/n$. 则对任何 $u>0$，有

$$P\left(\left|\sum_{i=1}^n \xi_i/n\right| \geqslant u\right) \leqslant 2\exp(-nu^2/(2\sigma^2 + 2bu)). \quad (3.14)$$

其证明参见文献[18].

现在转到定理 3.1 的证明：不失普遍性，不妨设 $\beta_0 = 0$，令

$$X_{ni} = S_n^{-1/2} x_i \quad (1 \leqslant i \leqslant n), \quad \beta_{n0} = S_n^{1/2} \beta_0. \quad (3.15)$$

而将模型(1.1)改写为

$$Y_i = x_{ni}'\beta_{n0} + e_i \quad (1 \leqslant i \leqslant n). \tag{3.16}$$

由 $\beta_0=0$ 得 $\beta_{n0}=0$. 又注意：

$$\sum_{i=1}^{n} x_{ni}x_{ni}' = I_p,\ \sum_{i=1}^{n} \| x_{ni} \|^2 = p,\ d_n = \max_{1\leqslant i\leqslant n} \| x_{ni} \|^2. \tag{3.17}$$

以 β_n^* 记在模型(3.16)下并使用函数 ρ 时，β_{n0}的 M 估计，则易见 β_n^* 与 $\hat{\beta}_n$ 有关系 $\hat{\beta}_n=S_n^{-1/2}\beta_n^*$.

取 $\varepsilon=l_1/2, 0<\eta<\min(1,\Delta,l_2), \tilde{b}_n=[\eta/\sqrt{d_n p}]$ （即$\eta/\sqrt{d_n p}$的整数部分），记

$$D_n = \{\beta = (\beta_1,\cdots,\beta_p)' : -\tilde{b}_n \leqslant \beta_i \leqslant \tilde{b}_n,\ 1 \leqslant i \leqslant p\}.$$

把超立方体 D_n 的各边分为 $2\tilde{b}_n^{m+1}$ 等分，从而把 D_n 剖分为$(2\tilde{b}_n^{m+1})^p$个超立方体，记为$\{B_j, 1\leqslant j\leqslant N_n\}$，$N_n=(2\tilde{b}_n^{m+1})^p$，$m$ 的值待定，B_j 的边长为 $\tilde{b}_n^{-m}$，其中心记为 b_j. 引进函数：

$$\begin{aligned}
\Phi_{ni}(\beta) &= \rho(e_i) - \rho(e_i - x_{ni}'\beta),\\
\Lambda_{ni}(\beta) &= E\Phi_{ni}(\beta),\\
R_{ni}(\beta) &= \Phi_{ni}(\beta) - \Lambda_{ni}(\beta).
\end{aligned}$$

定义事件：

$$E_n = \left\{\sup\left(\left|\sum_{i=1}^{n} R_{ni}(\beta)\right| : \beta \in D_n\right) \geqslant \varepsilon\tilde{b}_n^2\right\},$$

$$E_{n1} = \left\{\sup\left(\left|\sum_{i=1}^{n} R_{ni}(b_j)\right| : 1 \leqslant j \leqslant N_n\right) \geqslant b_n^2\varepsilon/3\right\},$$

$$E_{n2} = \left\{\sup(\sup\left(\left|\sum_{i=1}^{n} (R_{ni}(\beta) - R_{ni}(b_j))\right| : \beta \in B_j) : 1 \leqslant j \leqslant N_n\right) \geqslant \tilde{b}_n^2\varepsilon/3\right\}.$$

并分别以 p_n、p_{n1}和 p_{n2}记其概率，则易见：

$$p_n \leqslant p_{n1} + p_{n2}. \tag{3.18}$$

记 $g(x)=\max(|\Psi_+(x+\Delta)|, |\Psi_+(x-\Delta)|)$. 由(3.8)，有

$$Eg^m(e_1) \leqslant 2h_m \quad (m = 1,2,\cdots). \tag{3.19}$$

由 ρ 的凸性，知当 a、b 都落在区间$(x-\Delta,x+\Delta)$内时，有$|\rho(b)-\rho(a)|\leqslant|b-a|g(x)$. 当 $\beta\in D_n$ 时，有

$$|x_{ni}'b_j|\leqslant\sqrt{d_n p}\,\tilde{b}_n\leqslant\eta\leqslant\Delta,\quad |x_{ni}'\beta|\leqslant\sqrt{d_n p}\,\tilde{b}_n\leqslant\eta\leqslant\Delta.$$

于是当 $\beta\in B_j$ 时，有

$$\begin{aligned}|\rho(e_i-x_{ni}'b_j)-\rho(e_i-x_{ni}'\beta)|&\leqslant g(e_i)|x_{ni}'(\beta-b_j)|\\&\leqslant g(e_i)\sqrt{d_n}\,\|\beta-b_j\|\\&\leqslant\sqrt{p}\,g(e_i)\tilde{b}_n^{-m}\quad(1\leqslant j\leqslant N_n),\end{aligned}$$

故由(3.19)，得

$$\begin{aligned}&\sup\left\{\left|\sum_{i=1}^{n}(R_{ni}(\beta)-R_{ni}(b_j))\right|:\beta\in B_j\right\}\\&\quad\leqslant\sqrt{p}\,\tilde{b}_n^{-m}\sum_{i=1}^{n}g(e_i)+E\left(\sqrt{p}\,\tilde{b}_n^{-m}\sum_{i=1}^{n}g(e_i)\right)\\&\quad\leqslant\sqrt{p}\,\tilde{b}_n^{-m}\sum_{i=1}^{n}g(e_i)+n\sqrt{p}\,2h_1\tilde{b}_n^{-m}.\end{aligned}$$

由此得出

$$p_{n2}\leqslant N_nP\left(\sqrt{p}\,\tilde{b}_n^{-m}\sum_{i=1}^{n}g(e_i)+2h_1n\sqrt{p}\,\tilde{b}_n^{-m}\geqslant\tilde{b}_n^2\varepsilon/3\right).$$

根据条件(3.9)及 $\tilde{b}_n$ 的定义知，若取 m 充分大，将有

$$2h_1n\sqrt{p}\,\tilde{b}_n^{-m}<\tilde{b}_n^2\varepsilon/6,$$

因此对这样的 m，有

$$\begin{aligned}p_{n2}&\leqslant N_nP\left(\sum_{i=1}^{n}g(e_i)\geqslant\tilde{b}_n^{m+2}\varepsilon/6\sqrt{p}\right)\\&\leqslant N_n\left(\tilde{b}_n^{m+2}\varepsilon/6\sqrt{p}\right)^{-p-1}E\left(\sum_{i=1}^{n}g(e_i)\right)^{p+1}.\end{aligned}$$

根据(3.19)及 $N_n=(2\tilde{b}_n^{m+1})^p$，得

$$\begin{aligned}p_{n2}&\leqslant(2\tilde{b}_n^{m+1})^p(\tilde{b}_n^{m+2})^{-(p+1)}(\varepsilon/6\sqrt{p})^{-(p+1)}2h_{p+1}n^{p+1}\\&=O(\tilde{b}_n^{-(m+p+2)}n^{p+1}),\end{aligned}$$

由上式并根据(3.9)及 $\bar{b}_n$ 的定义知，若取 m 充分大时，有

$$p_{n2} = O(n^{-2}). \tag{3.20}$$

考虑 p_{n1}，与前面类似的推理，得出

$$|R_{ni}(b_j)| \leqslant (g(e_i) + 2h_1)|x_{ni}'b_j|.$$

因为 $|x_{ni}'b_j| \leqslant \sqrt{d_n}\bar{b}_n \leqslant \eta < 1$，根据(3.19)知，当 $t \geqslant 2$ 时，有

$$E|R_{ni}(b_j)|^t \leqslant K_t b_j' x_{ni} x_{ni}' b_j.$$

这里 K_t 只与 t 有关. 由此并注意(3.17)，有

$$E\sum_{i=1}^{n}|R_{ni}(b_j)|^t \leqslant K_t\|b_j\|^2 \leqslant pK_t\bar{b}_n^2. \tag{3.21}$$

特别是，取 $t=2$，得到

$$\sum_{i=1}^{n}\mathrm{Var}(R_{ni}(b_j)) \leqslant pK_2\bar{b}_n^2. \tag{3.22}$$

将引理 3.1 用于 $\xi_i = R_{ni}(b_j)$ $(1 \leqslant i \leqslant n)$，并取(3.12)式中的 $u = \bar{b}_n^2\varepsilon/3$，利用(3.21)和(3.22)，得到

$$\begin{aligned} P\left(\left|\sum_{i=1}^{n}R_{ni}(b_j)\right| \geqslant \bar{b}_n^2\varepsilon/3\right) & \\ \leqslant c_{1t}3^t\varepsilon^{-t}\bar{b}_n^{-2t}pK_t\bar{b}_n^2 &+ 2\exp(-c_{2t}9^{-1}\varepsilon^2\bar{b}_n^4/pK_2\bar{b}_n^2) \\ = \tilde{c}_{1t}\bar{b}_n^{-(2t-2)} &+ 2\exp(-\tilde{c}_{2t}\bar{b}_n^2). \end{aligned}$$

此处 $\tilde{c}_{1t} > 0$，$\tilde{c}_{2t} > 0$，都只与 t 有关. 由此，并注意 $N_n = (2\bar{b}_n^{m+1})^p$，得到

$$p_{n1} \leqslant (2\bar{b}_n^{m+1})^p(\tilde{c}_{1t}\bar{b}_n^{-(2t-2)} + 2\exp(-\tilde{c}_{2t}\bar{b}_n^2)). \tag{3.23}$$

根据(3.9)可知，只须取 t 足够大，就有 $p_{n1} = O(n^{-2})$. 将此与(3.18)、(3.20)结合起来并利用 Borel-Cantelli 引理，得到

$$P(E_n, \mathrm{i.o.}) = 0. \tag{3.24}$$

注意：此式对任意的满足前述条件的 η 都成立.

现定义事件：

$$\begin{aligned} F_n &= \left\{\sum_{i=1}^{n}\rho(Y_i) \geqslant \inf_{|\beta|=\bar{b}_n}\rho(Y_i - x_{ni}'\beta)\right\} \\ &= \left\{\sum_{i=1}^{n}\rho(e_i) \geqslant \inf_{|\beta|=\bar{b}_n}\rho(e_i - x_{ni}'\beta)\right\}. \end{aligned}$$

此处对任一向量 a，用 $|a|$ 表示 a 的各分量绝对值中最大者，又利用了 $\beta_0=0$，因而 $Y_i=e_i$. 现在进行证 $F_n\subset E_n$. 因为当 $\beta\in D_n$ 时，有 $|x_{ni}'\beta|\leqslant\eta<l_2$. 故依(3.7)，有

$$\Lambda_{ni}(\beta)\leqslant -l_1|x_{ni}'\beta|^2\quad(\text{当 }\beta\in D_n\text{ 时}).$$

由此，利用(3.17)，得到

$$\sum_{i=1}^{n}\Lambda_{ni}(\beta)\leqslant -l_1\sum_{i=1}^{n}|x_{ni}'\beta|^2=-l_1\|\beta\|^2\quad(\beta\in D_n).\tag{3.25}$$

由(3.25)，并注意到 $R_n(\beta)$ 的定义知，当事件 F_n 发生时，可找到 $\bar\beta$，$|\bar\beta|=\bar b_n$，使

$$\sum_{i=1}^{n}R_{ni}(\bar\beta)\geqslant l_1\bar b_n^2=2\varepsilon\bar b_n^2,$$

故有

$$\sup\left\{\left|\sum_{i=1}^{n}R_{ni}(\beta)\right|:\beta\in D_n\right\}\geqslant\varepsilon\bar b_n^2.$$

这就证明了 $F_n\subset E_n$. 由此及(3.24)，得到

$$P(F_n,\text{i.o.})=0.\tag{3.26}$$

由此，利用 ρ 的凸性及定理1.3的3°，得到

$$P(|\beta_n^*|>\bar b_n,\text{i.o.})=0.\tag{3.27}$$

根据(3.11)，存在常数 $h>0$，使 $d_n\geqslant h/\lambda_n$，λ_n 为 S_n 的最小特征根. 由此及 $\sqrt{d_n}\leqslant\eta/\bar b_n$ 可知，对任给的 $\varepsilon_0>0$，有

$$\begin{aligned}\{\|\hat\beta_n\|\geqslant\varepsilon_0\}&\subset\{\|S_n^{1/2}\hat\beta_n\|\geqslant\sqrt{\lambda_n}\varepsilon_0\}\\&\subset\{\|\beta_n^*\|\geqslant\varepsilon_0\sqrt{h}/\sqrt{d_n}\}\\&\subset\{\|\beta_n^*\|\geqslant\varepsilon_0\eta^{-1}\sqrt{h}\,\bar b_n\}\\&\subset\{|\beta_n^*|\geqslant p^{-1/2}\varepsilon_0\eta^{-1}\sqrt{h}\,\bar b_n\}.\end{aligned}$$

取 $\eta>0$充分小，使 $\eta<\min(1,\Delta,l_2)$且 $p^{-1/2}\varepsilon_0\eta^{-1}\sqrt{h}>1$，则由上式得 $\{\|\hat\beta_n\|\geqslant\varepsilon_0\}\subset\{|\beta_n^*|>\bar b_n\}$. 由此及(3.27)，得到

$$P(\|\hat\beta_n\|\geqslant\varepsilon_0,\text{i.o.})=0.$$

由于此式对任何 $\varepsilon_0>0$成立，知 $\hat{\beta}_n\to 0=\beta_0$，a. s.. 定理证毕. ▌

根据这个定理，若 Ψ_+为有界，则当(3.7)和(3.9)成立时，β_0 的 M 估计为强相合. 下一个定理表明，在此情况下，条件(3.9)还可以再加以改善.

定理3.2　设 $e_1,e_2,\cdots$为 iid.，ρ 为凸函数，满足(3.7)，且满足下述两条件之一：

1° 存在常数 $M<\infty$，使 $P(|e_1|<M)=1$；

2° Ψ_+有界，或等价地，ρ 满足 Lipshitz 条件.

则当

$$d_n = o(1/\log n) \tag{3.28}$$

时，$\hat{\beta}_n$ 为 β_0的强相合估计.

证　虽说由(3.9)到(3.28)的改进似乎很有限，但这一改进却是本质的，因而证明也是繁复的. 基本思想仍是上一定理证明中所用的分割法，但分割不是一次，而是需要多次.

仍将模型转化为(3.16)并设 $\beta_{n0}=\beta_0=0$，定义 η、$\bar{b}_n$ 和 D_n 如前，把 D_n 的各边分为 $\bar{b}_n^2$ 等分，从而将 D_n 分割成了 $\bar{b}_n^{2p}$ 个相等的超立方体：

$$\{B_{j_1}: 1\leqslant j_1\leqslant \bar{b}_n^{2p}\}.$$

接着，把每个 B_{j_1}的各边分为 $\bar{b}_n^2$ 等分，从而将 B_{j_1}分割成了 $\bar{b}_n^{2p}$ 个相等的超立方体：

$$\{B_{j_1j_2}: 1\leqslant j_2\leqslant \bar{b}_n^{2p}\}, 1\leqslant j_1\leqslant \bar{b}_n^{2p}.$$

一般地，在得到

$$\{B_{j_1\cdots j_{m-1}}: 1\leqslant j_1\leqslant \bar{b}_n^{2p},\cdots,1\leqslant j_{m-1}\leqslant \bar{b}_n^{2p}\}$$

后，把 $B_{j_1\cdots j_{m-1}}$的各边都分为 $\bar{b}_n^2$ 等分，从而将 $B_{j_1\cdots j_{m-1}}$分割成了 $\bar{b}_n^2$ 个相等的超立方体 $B_{j_1\cdots j_{m-1}j_m}: 1\leqslant j_m\leqslant \bar{b}_n^{2p}$，记为

$$G_m = \{B_{j_1\cdots j_m}: 1\leqslant j_1\leqslant \bar{b}_n^{2p},\cdots,1\leqslant j_m\leqslant \bar{b}_n^{2p}\},$$

G_m 中各立方体的边长都等于$2\bar{b}_n/\bar{b}_n^{2m}$. 因为 $\bar{b}_n\to\infty$，故知当 n 充分大时，$B_{j_1\cdots j_m}$的直径不超过 $\bar{b}_n^{-(2m-2)}$. 把 $B_{j_1\cdots j_m}$的中心记为 $b_{j_1\cdots j_m}$. 有时，我们分别以 B 和 b 记一个流动的 $B_{j_1\cdots j_m}$和 $b_{j_1\cdots j_m}$.

取 $\varepsilon \in (0, l_1/2)$，定义以下的量：

$$U_0 = \sum\nolimits_{B\in G_1} P\left(\left|\sum_{i=1}^{n} R_{ni}(b)\right| \geqslant \varepsilon \tilde{b}_n^2\right),$$

$$U_m = \sum_{j_1,\cdots,j_{m+1}=1}^{\tilde{b}_n^2} P\left(\left|\sum_{i=1}^{n} (R_{ni}(b_{j_1\cdots j_m}) - R_{ni}(b_{j_1\cdots j_{m+1}}))\right| \geqslant \varepsilon \tilde{b}_n^2/3^m\right) \quad (m \geqslant 1),$$

$$Q_m = \sum\nolimits_{B\in G_m} P\left(\sup_{\beta\in B}\left|\sum_{i=1}^{n} (R_{ni}(\beta) - R_{ni}(b))\right| \geqslant \varepsilon \tilde{b}_n^2/3^{m-1}\right) \quad (m \geqslant 1),$$

则易见有 $Q_m \leqslant U_m + Q_{m+1}$ $(m\geqslant 1)$，且

$$V \equiv P\left(\sup_{\beta\in D_n}\left|\sum_{i=1}^{n} R_{ni}(\beta)\right| \geqslant 2\varepsilon \tilde{b}_n^2\right) \leqslant U_0 + Q_1.$$

于是，对任何自然数 N，有

$$V \leqslant \sum_{m=0}^{N} U_m + Q_{N+1}. \tag{3.29}$$

考察

$$\Phi_{ni}(\beta) - \Phi_{ni}(b_{j_1\cdots j_m}) \quad (\beta \in B_{j_1\cdots j_m}),$$

如果定理3.2的条件1°满足，则令

$$H_1 = \sup\{|\Psi_+(u)| : |u| \leqslant M + 1\},$$

如果条件2°满足，则令

$$H_2 = \sup\{|\Psi_+(u)| : |u| < \infty\},$$

令

$$H = \max(H_1, H_2).$$

考虑到当 $\beta\in D_n$ 时，$|x_{ni}'\beta| \leqslant \eta \leqslant 1$，因而 $|e_i - x_{ni}'\beta| \leqslant M+1$. 将有

$$|\Phi_{ni}(\beta) - \Phi_{ni}(b_{j_1\cdots j_m})| \leqslant H|x_{ni}'(\beta - b_{j_1\cdots j_m})| \leqslant H\cdot\|x_{ni}\|\cdot\|\beta - b_{j_1\cdots j_m}\|.$$

因为 $\|\beta - b_{j_1\cdots j_m}\|$ 当 $\beta\in B_{j_1\cdots j_m}$ 时，不超过 $B_{j_1\cdots j_m}$ 的直径，故不超过 $\tilde{b}_n^{-(2m-2)}$，而

$$\sum_{i=1}^{n}\|x_{ni}\| \leqslant \left(n\sum_{i=1}^{n}\|x_{ni}\|^2\right)^{1/2} = \sqrt{np},$$

有

$$\left|\sum_{i=1}^{n}(\Phi_{ni}(\beta)-\Phi_{ni}(b_{j_1\cdots j_m}))\right|\leqslant H\sqrt{np}\bar{b}_n^{-2(m-1)}\quad(\beta\in B_{j_1\cdots j_m}),$$

由此推出：

$$\left|\sum_{i=1}^{n}(R_{ni}(\beta)-R_{ni}(b_{j_1\cdots j_m}))\right|\leqslant 2H\sqrt{np}\bar{b}_n^{-2(m-1)}\quad(\beta\in B_{j_1\cdots j_m}).\tag{3.30}$$

因为$\lim\limits_{n\to\infty}\bar{b}_n=\infty$，知当 n 充分大时，可找到 M_n，当 $m>M_n$ 时，使

$$2H\sqrt{np}\bar{b}_n^{-2(m-1)}<\varepsilon\bar{b}_n^2/3^{m-1}.$$

由此，注意到 Q_m 的定义及(3.30)式，即知

$$Q_{M_n+1}=0.$$

再由(3.29)，即得

$$V\leqslant\sum_{m=0}^{M_n}U_m.\tag{3.31}$$

因为

$$|\Phi_{ni}(b_j)|\leqslant H\cdot|x_{ni}'b_j|\quad(1\leqslant i\leqslant n),$$

有

$$|R_{ni}(b_j)|\leqslant 2H\cdot|x_{ni}'b_j|\quad(1\leqslant i\leqslant n).$$

故若将引理3.2应用于$\sum\limits_{i=1}^{n}R_{ni}(b_j)$，则(3.13)式中的$\sum\limits_{i=1}^{n}(b_i-a_i)^2$应为

$$\sum_{i=1}^{n}(4Hx_{ni}'b_j)^2=16H^2\|b_j\|^2\leqslant 16H^2p\bar{b}_n^2.$$

于是由(3.13)，得到

$$P\left(\left|\sum_{i=1}^{n}R_{ni}(b_j)\right|\geqslant\varepsilon\bar{b}_n^2\right)\leqslant 2\exp(-2\varepsilon^2\bar{b}_n^4/16H^2p\bar{b}_n^2)=2\exp(-c\bar{b}_n^2),\tag{3.32}$$

此处取 $c=2\varepsilon^2/16H^2p$. 类似地，因为

$$|\Phi_{ni}(b_{j_1\cdots j_m})-\Phi_{ni}(b_{j_1\cdots j_{m+1}})|\leqslant H\cdot|x_{ni}'(b_{j_1\cdots j_m}-b_{j_1\cdots j_{m+1}})|,$$

而$\|b_{j_1\cdots j_m}-b_{j_1\cdots j_{m+1}}\|$不超过 $B_{j_1\cdots j_m}$的直径，故不超过 $\bar{b}_n^{-2(m-1)}$. 由此，

仿照得到(3.32)的推理,用引理3.2,得到

$$P\left(\left|\sum_{i=1}^{n}(R_{ni}(b_{j_1\cdots j_m})-R_{ni}(b_{j_1\cdots j_{m+1}}))\right|\geqslant\varepsilon\bar{b}_n^2/3^m\right)$$

$$\leqslant 2\exp(-2\varepsilon^2\bar{b}_n^4 3^{-2m}/16H^2p\bar{b}_n^{-4(m-1)})$$

$$=2\exp(-9^{-m}c\bar{b}_n^{4m}). \tag{3.33}$$

由 U_m 的定义,利用(3.32)和(3.33),并注意到(3.31),得到

$$V\leqslant 2\bar{b}_n^{2p}\exp(-c\bar{b}_n^2)+2\sum_{m=1}^{\infty}\bar{b}_n^{2(m+1)}\exp(-9^{-m}c\bar{b}_n^{4m}).$$

由于$\lim\limits_{n\to\infty}\bar{b}_n=\infty$,由上式易知:若取 $c_0\in(0,c)$,则当 n 充分大时,有

$$V\leqslant\exp(-c_0\bar{b}_n^2). \tag{3.34}$$

按假定(3.7),以及当 $\beta\in D_n$ 时,$|x_{ni}'\beta|\leqslant\eta<l_2$,得到

$$\sum_{i=1}^{n}\Lambda_{ni}(\beta)\leqslant -l_1\sum_{i=1}^{n}(x_{ni}'\beta)^2=l_1\|\beta\|^2.$$

由此及 V 的定义,并注意到(3.34),即得

$$P\left(\sup_{|\beta|=\bar{b}_n}\left(\sum_{i=1}^{n}(\rho(e_i)-\rho(e_i-x_{ni}'\beta))\right)>(2\varepsilon-l_1)\bar{b}_n^2\right)$$

$$\leqslant V\leqslant\exp(-c_0\bar{b}_n^2).$$

因为 $\varepsilon<l_1/2$,由上式得

$$P\left(\sum_{i=1}^{n}\rho(e_i)-\inf_{|\beta|=\bar{b}_n}\sum_{i=1}^{n}\rho(e_i-x_{ni}'\beta)>0\right)\leqslant\exp(-c_0\bar{b}_n^2). \tag{3.35}$$

因为 ρ 为凸函数,利用定理1.3的3°,根据(3.35),得

$$P(|\beta_n^*|>\bar{b}_n)\leqslant\exp(-c_0\bar{b}_n^2). \tag{3.36}$$

由(3.28),知$\sum\limits_{n=1}^{\infty}\exp(-c_0\bar{b}_n^2)<\infty$,故用 Borel-Cantelli 引理,仍得(3.27).于是根据(3.27)式以下的推理,得出

$$\hat{\beta}_n\to 0=\beta_0,\text{a.s.}.$$

从而完成了本定理的证明. ▌

注 1 ρ 为凸函数的假定的关键作用在于由(3.35)推出(3.36).在定理3.1的证明中其作用相同,而这是基于凸函数的一

个简单性质——定理1.3的3°. 如果 ρ 为非凸函数,则证明全盘失效. 在 M 估计的理论中,ρ 为凸函数的假定之所以有如此重要的地位,就由于这一简单性质.

注 2　定理3.1和3.2容易推广到 e_i 独立但不必为同分布的情况. 只须对假定作少许修改:对定理3.1,要假定(3.7)对任何 e_i 成立且 l_1、l_2与 i 无关,(3.8)式对任何 i 成立且 h_m 与 i 无关. 对定理3.2,假定(3.7)要作上述修改,且其假定1°对任何 i 成立且 M 与 i 无关.

注 3　定理3.2用于 LADE 的特例,推广了吴月华于1988年在文献[65]中作出的一项结果:设在线性模型(1.1)中,$e_1,e_2,\cdots$为 iid., $\mathrm{med}(e_1)=0$,且存在常数 $l_1>0$、$l_2>0$,使

当 $h\in(0,l_2)$时,$P(0<e_1<h)\geqslant l_1h\leqslant P(-h<e_1<0)$. 记

$t_n=\max\limits_{1\leqslant i\leqslant n}\|x_i\|$, $\lambda_n=S_n$ 的最小特征根.

则当

$$t_n^2/\lambda_n=o(1/\log n),$$

且存在 $a>0$,使 $t_n=O(n^a)$时,β_0的 LADE 是强相合的.

显然, $d_n\leqslant t_n^2/\lambda_n$, 故当 $t_n^2/\lambda_n=o(1/\log n)$时,必有

$$d_n = o(1/\log n).$$

这样,定理3.2(用于 LADE)使上述第一条件减弱了,且免去了第二条件. 但值得指出的是:吴月华的上述结果是 LADE 强相合问题的第一个结果,也是除 LSE 之外有关 M 估计的强收敛问题的第一个结果.

注 4　许宝騄教授和 Robbins 教授曾引进"完全收敛"的概念(参见文献[32]):称一串随机向量$\{\xi_n\}$完全收敛于 a,若对任给的 $\varepsilon>0$,有$\sum\limits_{n=1}^{\infty}P(\|\xi_n-a\|\geqslant\varepsilon)<\infty$. 按这个说法,在定理3.1和定理3.2的条件下,我们都证明了 β_0的 M 估计 $\hat{\beta}_n$ 完全收敛于 β_0.

由 ξ_n 完全收敛于 a 导出 $\xi_n\to a$, a. s.. 但其逆不必为真. 直到现在,尚未能举出强相合但并非完全收敛的 LADE 的例子来.

注 5　一个悬而未决的问题是:β_0的一分量,或一般地说,β_0的

一线性函数 $c'\beta_0$的 M 估计 $c'\hat{\beta}_n$ 强相合的条件. 这种条件当然应该比(3.9)或(3.28)弱一些. 对弱相合,类似的问题已由定理2.4解决了.

注6 对于一个特定的满足定理3.1的条件1°、2°或定理3.2中相应条件的误差序列$\{e_i\}$而言,无论(3.9)或是(3.28),都不是 $\hat{\beta}_n$ 强相合的必要条件.

考察线性模型

$$Y_i = x_i\beta_0 + e_i, \ x_i = 10^i \quad (1 \leqslant i \leqslant n),$$

其中 $e_1, e_2, \cdots$为 iid., 有公共分布 $R(-1/2, 1/2)$. 按本章开始处(§3.1前面)那个例子中的论证,在此模型中,β_0的 LADE $\hat{\beta}_n$ 为强相合. 但此处有

$$d_n = \frac{99}{100}(1 - 100^{-n})^{-1} \to \frac{99}{100},$$

故甚至 $d_n \to 0$的条件也不满足,更谈不到(3.9)或(3.28)了.

这个例子中 e_i 的分布是给定了,在本章初的讨论中,我们已看到:若 e_i 的分布是另外一个样子,则 $\hat{\beta}_n$ 就不为强相合了. 这显示,随着 e_i 分布的不同,对 d_n 的要求也会有不同,这样就产生一个问题:针对一切满足定理3.1或3.2的条件的分布,定理中对 d_n 的要求还有无可能降低?比如说,把(3.9)减弱为(3.28),定理3.1的结论是否仍成立?或者,把(3.28)减弱为 $d_n = 0(1/\log\log n)$,定理3.2的结论是否仍成立?对一般的凸函数 ρ,这个问题还不清楚. 但对重要的 LADE,已有基本的但尚未完全彻底的结论,这个十分细致而困难的问题将在下节加以讨论.

模型带常数项的情况

考虑带常数项 α_0的线性模型:

$$Y_i = \alpha_0 + x_i'\beta_0 + e_i \quad (1 \leqslant i \leqslant n), \tag{3.37}$$

这个模型是(1.1)的特例,自然可以把定理3.1和定理3.2直接应用于它,而得出 α_0和 β_0的 M 估计 $\hat{\alpha}_n$ 和 $\hat{\beta}_n$ 强相合的条件. 但是,如同弱相合的情况一样(参见§2.1中第四段),针对(3.37)的特殊形式,条件(3.9)或(3.28)可写成另外的形式,只是涉及低一阶的方

阵.

记 $\bar{x}_n=\sum_{i=1}^{n} x_i/n$，$\alpha_n=\alpha_0+\bar{x}_n'\beta_0$，把模型(3.37)改写为

$$Y_i=\alpha_n+z_{ni}'\beta_0+e_i \quad (1\leqslant i\leqslant n), \tag{3.38}$$

记 $T_n=\sum_{i=1}^{n} z_{ni}z_{ni}'=\sum_{i=1}^{n}(x_i-\bar{x}_n)(x_i-\bar{x}_n)'$，而

$$\delta_n=\max_{1\leqslant i\leqslant n}(x_i-\bar{x}_n)'T_n^{-1}(x_i-\bar{x}_n). \tag{3.39}$$

定理3.3　设在模型(3.37)中 e_i 为iid.，且连同函数 ρ 一起，满足定理3.1或3.2的条件，则当 $\delta_n=O(n^{-\delta})$（对某个 $\delta>0$）或 $\delta_n=o(1/\log n)$ 时，α_0、β_0 的 M 估计 $\hat{\alpha}_n$ 和 $\hat{\beta}_n$ 都是强相合的.

证　对模型(3.38)来说，S_n 矩阵有形式：

$$\tilde{S}_n=\begin{pmatrix} n & 0 \\ 0 & T_n \end{pmatrix}.$$

于是此模型的 d_n 值为

$$\begin{aligned}\tilde{d}_n&=\max_{1\leqslant i\leqslant n}(1,(x_i-\bar{x}_n)')\tilde{S}_n^{-1}(1,(x_i-\bar{x}_n)')'\\ &=1/n+\delta_n,\end{aligned}$$

其中 δ_n 见(3.39). 因此，若 $\delta_n=O(n^{-\delta})$ 或 $\delta_n=O(1/\log n)$，则 $\tilde{d}_n$ 也如此. 因而按定理3.1或3.2，α_n 及 β_0 的 M 估计即 $\hat{\alpha}_{n0}=\hat{\alpha}_n+\bar{x}_n'\hat{\beta}_n$ 及 $\hat{\beta}_n$，分别是 α_n 及 β_0 的强相合估计，此处 $\hat{\alpha}_n$ 和 $\hat{\beta}_n$ 分别是在模型(3.37)之下 α_0 和 β_0 的 M 估计，由此立即得出 $\hat{\beta}_n$ 的强相合性，为证 $\hat{\alpha}_n$ 的强相合性，不失普遍性，不妨设 $\alpha_0=\beta_0=0$. 因为

$$\hat{\alpha}_{n0}\equiv\hat{\alpha}_n+\bar{x}_n'\hat{\beta}_n,\ \alpha_n\equiv\alpha_0+\bar{x}_n'\beta_0,\ \hat{\alpha}_{n0}-\alpha_n\to 0,\ \text{a.s.},$$

可知欲证　$\hat{\alpha}_n\to\alpha_0=0$, a.s.,

只须证　$\bar{x}_n'\hat{\beta}_n\to 0$, a.s..

以 λ_n 记 T_n 的最小特征根，则依(3.11)，存在常数 $h>0$，使 $\delta_n\geqslant h/\lambda_n$. 所以

$$x_1'T_n^{-1}x_1\leqslant\|x_1\|^2/\lambda_n\leqslant\|x_1\|^2\delta_n/h\equiv c\delta_n,\ c=\|x_1\|^2/h.$$

又按 δ_n 的定义，有 $(x_1-\bar{x}_n)'T_n^{-1}(x_1-\bar{x}_n)\leqslant\delta_n$，故

$$\bar{x}_n'T_n^{-1}\bar{x}_n\leqslant 2x_1'T_n^{-1}x_1+2(x_1-\bar{x}_n)'T_n^{-1}(x_1-\bar{x}_n)$$

$$\leqslant 2c\delta_n + 2\delta_n \equiv c'\delta_n,\ c' = 2c + 2. \tag{3.40}$$

前已算得模型(3.38)的 S_n 矩阵与 d_n 的值分别为 $\tilde{S}_n$ 与 $\tilde{d}_n=1/n+\delta_n\leqslant 2\delta_n$(因为 $\delta_n\geqslant p/n$). 按定理3.1及定理3.2证明过程中所得(例如(3.36)式),对充分小的 $\eta>0$,有

$$P\left(\left|\begin{pmatrix} n^{-1/2} & 0 \\ 0 & T_n^{-1/2}\end{pmatrix}\begin{pmatrix}\hat{\alpha}_{n0} \\ \hat{\beta}_n\end{pmatrix}\right| \geqslant \eta\Big/\sqrt{\tilde{d}_n}\right)\leqslant \exp\left(-c_0\left[\eta\Big/\sqrt{\tilde{d}_n}\right]^2\right).$$

此处仍如前,以 $|a|$ 记向量 a 的各分量绝对值最大者. 由于 $\delta_n\leqslant\tilde{d}_n\leqslant 2\delta_n$ 而 $\delta_n\to 0$,由上式得

$$P\left(|T_n^{-1/2}\hat{\beta}_n|\geqslant \eta/\sqrt{\delta_n}\right)\leqslant \exp(-c_0'/\delta_n). \tag{3.41}$$

此处 c_0'是一个与 n 无关的正常数. 因为按(3.40),有

$$|\bar{x}_n'\hat{\beta}_n| = |\bar{x}_n'T_n^{-1/2}T_n^{1/2}\hat{\beta}_n| \leqslant (\bar{x}_n'T_n^{-1}\bar{x}_n)^{1/2}\|T_n^{1/2}\hat{\beta}_n\|$$

$$\leqslant \sqrt{c'\delta_n}\sqrt{p}\,|T_n^{1/2}\hat{\beta}_n|,$$

由此及(3.41)得知,若记 $\eta=\varepsilon/\sqrt{c'p}$,则

$$P(|\bar{x}_n'\hat{\beta}_n|\geqslant\varepsilon)\leqslant P(|T_n^{-1/2}\hat{\beta}_n|\geqslant\eta/\sqrt{\delta_n})\leqslant\exp(-c_0'/\delta_n).$$

按对 δ_n 的假定,知 $\sum_{n=1}^{\infty}\exp(-c_0'\delta_n)<\infty$,对任何 $c_0'>0$,由此得出 $\bar{x}_n'\hat{\beta}_n\to 0$, a.s.. 得所欲证. ▎

§3.2 LADE强相合条件可否改进

在 ρ 和 e_i 的公共分布适合一定的假设前提下,为使 β_0的 M 估计为强相合,要求$\{x_i\}$必须满足什么条件?在§3.1中,我们曾见到即使对弱相合,这个问题也解决得很不彻底:除了对 LSE 有比较完整的解决外,只是对 LADE 有了部分的解决,而对一般的凸函数 ρ,则只有很初步的结果,对强相合来说,问题解决得更差了:也只是在 LSE 的情况下,黎子良等在1979年的工作(见文献[44])得

出了基本的结果. 他们证明了：在误差 $e_1, e_2, \cdots$ 为 iid., $Ee_1=0$ 且 $0<Ee_1^2<\infty$（条件还可以适当放宽，见文献[44]）的假定下，$S_n^{-1}\to 0$ 是 β_0 的 LSE 强相合的充要条件. 此处问题的要害在于条件 $S_n^{-1}\to 0$ 的充分性. 因为前此已知（见文献[27]），对 LSE 弱相合而言，$S_n^{-1}\to 0$ 为必要的，因而对强相合来说，当然也为必要的. 但是否强相合的必要条件更严些？在黎子良等的结果之前并不明显. 现在既已证明了 $S_n^{-1}\to 0$ 对强相合的充分性，它当然也就是无可改进的强相合必要条件了.

除 LSE 外，对 M 估计强收敛的必要条件问题，迄今知之甚少. 这个问题难度之大，使我们难以期望它可轻易地获得较彻底的解决. 本节要证明一个结果，从一定角度看它，有些必要条件的意味，但还不能说是通常意义下的必要条件.

定理3.4 设在模型(1.1)中 $e_1, e_2, \cdots$ 为 iid., $\text{med}(e_1)=0$, e_1 在0点的一个领域内有概率密度 f, f 在0点连续，且 $f(0)>0$. 则对任何常数串 $D_n\uparrow\infty$ 来说，条件

$$d_n=o(D_n/\log n) \tag{3.42}$$

不再成为 β_0 的 LADE 强相合的充分条件.

这个结果的含义，从充分性的角度是易于解释的. 前已证明了 $d_n=o(1/\log n)$ 对 LADE $\hat{\beta}_n$ 强相合是充分的. 自然要问：能否把这条件再减轻些？即能否以一个趋于 ∞ 较 $\log n$ 为慢的函数 $a(n)$ 来取代 $\log n$，即找到 $a(n)\uparrow\infty$, $\log n/a(n)\to\infty$，但 $d_n=o(1/a(n))$ 仍是 $\hat{\beta}_n$ 强相合的充分条件？本定理的结论说明，这在"总体上"不可能. 就是说，如果 $a(n)$ 满足上述条件，则必可找到模型(1.1)，其中 e_1, $e_2, \cdots$ 满足定理3.4中的条件，而 $d_n=o(1/a(n))$，但 $\hat{\beta}_n$ 不为强相合. 这并不是说，对于任一个特定的满足定理3.4条件的模型(1.1)，条件 $d_n=o(1/\log n)$ 一定不可缺. 例如对 $p=1$ 的模型 $Y_i=x_i\beta_0+e_i$，若 e_i 为 iid.，公共分布为 $R(-1,1)$，则只要 $\lim\limits_{i\to\infty}|x_i|=\infty$，必有 $\hat{\beta}_n$ 为强相合（见 § 3.1之前的那段论证）.

定理3.4的意义在于：若把 d_n 趋于0的速度作为一个尺度，则

$1/\log n$ 这个数量级是一个"门槛",高于这个"门槛"者能保证强相合性;而低于它的则不行.

下面转到定理的证明,先证明几条引理.定义一串常数$\{C_n\}$如下:不妨设 $D_1=1$.令

$$C_1=\sqrt{D_1},\ C_n=\min(C_{n-1}+(n\log n)^{-1},\sqrt{D_n})\quad(n\geqslant 2). \tag{3.43}$$

引理3.4 序列$\{C_n\}$有以下的性质:

$$C_n=o(D_n), \tag{3.44}$$

$$1=C_1\leqslant C_2\leqslant\cdots\uparrow\infty, \tag{3.45}$$

$$nC_n/\log n\uparrow\infty\quad(n\geqslant 3), \tag{3.46}$$

$$C_n/\log n\downarrow 0, \tag{3.47}$$

$$\begin{aligned}&nC_n/\log n-(n-1)C_{n-1}/\log(n-1)\\&\geqslant C_n(\log(n-1)-1)/(\log n\cdot\log(n-1))\quad(n\geqslant 3),\end{aligned} \tag{3.48}$$

$$\begin{aligned}&nC_n/\log n-(n-1)C_{n-1}/\log(n-1)\\&\leqslant C_n/\log n\quad(n\geqslant 3),\end{aligned} \tag{3.49}$$

$$\exp(Ct/\log t)/\sqrt{\log t}\uparrow\quad(\text{当 } t \text{ 充分大}), \tag{3.50}$$

式中$C>0$为常数.

证 由(3.43)和$C_n\leqslant\sqrt{D_n}=o(D_n)$,即(3.43).为证(3.45),注意,若$C_{n-1}+(n\log n)^{-1}\leqslant\sqrt{D_n}$,则$C_n=C_{n-1}+(n\log n)^{-1}>C_{n-1}$.若不然,则$C_n=\sqrt{D_n}\geqslant\sqrt{D_{n-1}}\geqslant C_{n-1}$.故总有$C_{n-1}\leqslant C_n$.因此$a=\lim\limits_{n\to\infty}C_n\leqslant\infty$存在.若$a<\infty$,则因为$D_n\to\infty$,当$n$甚大时,将有$C_{n-1}+(n\log n)^{-1}<\sqrt{D_n}$,故当$n$甚大时,有$C_n=C_{n-1}+(n\log n)^{-1}$.而这将导致

$$C_n=C_m+\sum_{i=m+1}^{n}(i\log i)^{-1}\to\infty,\text{当 } n\to\infty.$$

这与$a<\infty$相矛盾.故必有$a=\infty$,这证明了(3.45).而(3.46)是显然的,因为$C_n\uparrow$,而$n\geqslant 3$时也有$n/\log n\uparrow$.为证(3.47),注意,按(3.42)有$C_n\leqslant C_{n-1}+(n\log n)^{-1}$,因而

$$C_n \leqslant 1 + \sum_{i=2}^{n} (i\log i)^{-1} = O(\log \log n).$$

这证明了　　　　　　$C_n/\log n \to 0.$

又因为 $\log n - \log(n-1) \geqslant n^{-1}$，故有

$$\begin{aligned}
&C_{n-1}/\log(n-1) - C_n/\log n \\
&\geqslant C_{n-1}/\log(n-1) - (C_{n-1} + (n\log n)^{-1})/\log n \\
&= C_{n-1}(\log n - \log(n-1))/(\log n \cdot \log(n-1)) - (n\log^2 n)^{-1} \\
&\geqslant (n\log n \cdot \log(n-1))^{-1} - (n\log^2 n)^{-1} > 0 \quad (n \geqslant 3).
\end{aligned}$$

这证明了 $C_n/\log n \downarrow$. 因而证明了(3.47)(此处还利用了 $C_{n-1} \geqslant C_1 = 1$). (3.49)容易由(3.47)推出来：

$$\begin{aligned}
&nC_n/\log n - (n-1)C_{n-1}/\log(n-1) \\
&\quad = C_n/\log n - (n-1)(C_{n-1}/\log(n-1) - C_n/\log n) \\
&\quad \leqslant C_n/\log n.
\end{aligned}$$

为证(3.48)，由上面的等式出发，并注意

$$\log n - \log(n-1) \leqslant (n-1)^{-1},$$

有

$$\begin{aligned}
&nC_n/\log n - (n-1)C_{n-1}/\log(n-1) \\
&\quad \geqslant C_n/\log n - (n-1)(C_{n-1}/\log(n-1) - C_{n-1}/\log n) \\
&\quad = C_n/\log n - (n-1)C_{n-1}(\log n - \log(n-1))/ \\
&\qquad (\log n \cdot \log(n-1)) \\
&\quad \geqslant C_n/\log n - C_{n-1}/(\log n \cdot \log(n-1)) \\
&\quad \geqslant C_n/\log n - C_n/(\log n \cdot \log(n-1)) \\
&\quad = C_n(\log(n-1) - 1)/(\log n \cdot \log(n-1)).
\end{aligned}$$

最后，(3.50)可以直接由对函数 $(\log t)^{-1/2}\exp(Ct/\log t)$ 求导而得出. 引理证毕. ▌

引理3.5　定义

$$\begin{aligned}
&x_1 = 1, x_i = (C_i/\log i)^{1/2}\exp(2^{-1}iC_i/\log i) \quad (i \geqslant 2), \\
&S_n = x_1^2 + \cdots + x_n^2,
\end{aligned} \tag{3.51}$$

则有

$$x_i \uparrow \infty, \text{当 } i \text{ 充分大}, \tag{3.52}$$

$$\max_{1\leq i\leq n} x_i^2/S_n = o(D_n/\log n). \qquad (3.53)$$

证 因为$(\log(n-1)/\log n)^{1/2}>1-2(n\log n)^{-1}$,当 n 充分大,有

$$x_n/x_{n-1}\geq (1-2(n\log n)^{-1})\exp(2^{-1}(nC_n/\log n - (n-1)C_{n-1}/\log(n-1))).$$

由(3.48),当 n 充分大时,有

$$nC_n/\log n - (n-1)C_{n-1}/\log(n-1) \geq 2^{-1}C_n/\log n \geq (2\log n)^{-1}.$$

故当 n 充分大时,有

$$x_n/x_{n-1}\geq (1-2(n\log n)^{-1})\exp((4\log n)^{-1}) \geq (1-2(n\log n)^{-1})(1+(4\log n)^{-1})>1,$$

因而证明了 $x_n\uparrow$,当 n 充分大.又因 $C_n\to\infty$, $x_n\to\infty$是显然的.这证明了(3.52).

利用(3.47),并记 $b_n=C_n/\log n$.依(3.46),$nb_n\uparrow\infty$.故

$$S_n\geq b_n\sum_{i=2}^{n}e^{ib_n} = b_n(e^{(n+1)b_n}-e^{2b_n})/(e^{b_n}-1) = (1+o(1))e^{(n+1)b_n} = (1+o(1))e^{nb_n}.$$

由此及(3.52),得知当 n 充分大时,有

$$\max_{1\leq i\leq n} x_i^2/S_n = x_n^2/S_n \leq (1+o(1))b_n\exp(nb_n-nb_n) = (1+o(1))b_n = o(D_n/\log n). \qquad (3.54)$$

最后一步用了(3.44).引理证毕. ▌

引理3.6 设$\{x_i\}$按(3.51)定义,记 $M(n)=\sum_{i=1}^{n}x_i$.则当 n 充分大时,有

$$M(n-[2/b_n]) < M(n)/2, \qquad (3.55)$$

此处如前 $b_n=C_n/\log n$.

证 利用(3.47),有

$$M(n)\geq \sqrt{b_n}\sum_{i=2}^{n}e^{ib_n/2} = \sqrt{b_n}(e^{(n+1)b_n/2}-e^{b_n})/(e^{b_n/2}-1) = (2+o(1))b_n^{-1/2}e^{nb_n/2}. \qquad (3.56)$$

另一方面，有

$$M(n) \leqslant 1 + \sum_{i=2}^{n} (C_n/\log i)^{1/2} \exp(2^{-1} i C_n/\log i)$$

$$= 1 + \sum_{i=2}^{[n/2]-1} + \sum_{i=[n/2]}^{n} = 1 + J_{1n} + J_{2n}. \qquad (3.57)$$

因为$1/\log i < 2$（当 $i \geqslant 2$），当 n 充分大时，有

$$J_{1n} \leqslant n\sqrt{C_n} \exp(nC_n/4\log(n/2-2))$$

$$\leqslant nC_n e^{nb_n/3} = o(e^{nb_n/2}). \qquad (3.58)$$

为估计 J_{2n}，令 $H(x = \int_2^x \exp(C_n t/2\log t)\mathrm{d}t$. 考虑到(3.50)，有

$$J_{2n} \leqslant \sqrt{C_n} \int_{[n/2]}^{n+1} (\log x)^{-1/2} \mathrm{d}H(x)$$

$$\leqslant \sqrt{C_n} H(n+1)/\sqrt{\log(n+1)}$$

$$+ \sqrt{C_n} \int_{[n/2]}^{n+1} H(x) \mathrm{d}(-(\log x)^{-1/2})$$

$$\leqslant \sqrt{C_n} H(n+1)/\sqrt{\log(n+1)}$$

$$+ \sqrt{C_n} H(n+1)(\log^{-1/2}[n/2] - \log^{-1/2}(n+1)).$$

$$(3.59)$$

因为当 $x \to \infty$时，有

$$\mathrm{d}(2^{-1}C_n x/\log x) = 2^{-1}C_n(\log x - 1)/\log^2 x$$

$$= 2^{-1}C_n(1+o(1))/\log x,$$

知当 $n \to \infty$时，有

$$H(n+1) = (2+o(1))C_n^{-1} \int_2^{n+1} (\log x)\mathrm{d}(\exp(2^{-1}C_n x/\log x))$$

$$= (2+o(1))C_n^{-1}\log(n+1)$$

$$\times \exp(2^{-1}C_n(n+1)/\log(n+1))$$

$$= (2+o(1))C_n^{-1}\log n \cdot \exp(2^{-1}nb_n). \qquad (3.60)$$

最后一式因为

$$0 \leqslant C_n((n+1)/\log(n+1) - n/\log n)$$

$$\leqslant C_n/\log n \to 0,$$

由(3.57)～(3.60)，得

$$M(n) \leqslant (2+o(1))C_n^{1/2}\log n/\sqrt{\log(n+1)}e^{nb_n/2}$$
$$= (2+o(1))b_n^{-1/2}e^{nb_n/2}. \tag{3.61}$$

将此式与(3.56)结合，得

$$M(n) = (2+o(1))b_n^{-1/2}e^{nb_n/2}. \tag{3.62}$$

由(3.47)和(3.62)，得

$$M(n-[2/b_n]) \leqslant (2+o(1))b_n^{-1/2}$$
$$\times \exp(2^{-1}(n-[2/b_n])C_n/\log(n-[2/b_n])). \tag{3.63}$$

因为

$$\lim_{n\to\infty}[2/b_n]C_n/\log(n-[2/b_n]) = 2, \tag{3.64}$$

$$nC_n/\log(n-[2/b_n]) = nb_n + O(1/\log n), \tag{3.65}$$

由(3.63)得知，当 n 充分大时，有

$$M(n-[2/b_n])$$
$$\leqslant (2+o(1))b_n^{-1/2}e^{nb_n/2}e^{-1} < M(n)/2. \tag{3.66}$$

这证明了(3.55)．从而引理证毕．▌

现在回到定理3.4的证明．先考虑模型(1.1)$p=1$的情况．按(3.51)定义$\{x_i\}$，并设$e_1,e_2,\cdots$为iid.，其公共分布定义如下：找充分大的自然数N，使$l\equiv e^{-C_N/4}<1/2$．令

$$P(e_1=x_n)=P(e_1=-x_n)$$
$$=e^{-C_n/4}-e^{-C_{n+1}/4} \quad (n\geqslant N), \tag{3.67}$$

而在区间$(-x_N,x_N)$内，e_1有密度$(1-2l)/(2x_N)$．这样定义的$\{e_i\}$显然满足定理3.4中的要求，并且根据引理3.5，条件(3.42)满足．但在下面我们要证明：β_0的 LADE $\hat{\beta}_n$ 不是强相合的．

不失普遍性，设$\beta_0=0$，故$Y_i=e_i$．对每个整数$k\geqslant2$，令$n_k=[k\log k]$，$m_k=n_k-[2/b_{n_k}]$．此处如前，记$b_n=C_n/\log n$．定义以下事件：

$$E_k=\{e_i\geqslant x_i,\ m_k\leqslant i\leqslant n_k\}, \tag{3.68}$$

$$\widetilde{E}_k = \{e_i \geqslant x_{n_k},\ m_k \leqslant i \leqslant n_k\}. \tag{3.69}$$

根据引理2.1和引理3.6,可知当 k 充分大而事件 E_k 发生时,有 $\hat{\beta}_{n_k} \geqslant 1$. 因此,如能证明

$$P(E_k, \text{i.o.}) = 1, \tag{3.70}$$

则 $P(\hat{\beta}_{n_k} \to 0) = 0$,而 $\hat{\beta}_n$ 不是强相合的.

为证(3.70),注意按(3.52)可知,当 k 充分大时,有 $E_k \supset \widetilde{E}_k$,故只须证明

$$P(\widetilde{E}_k, \text{i.o.}) = 1. \tag{3.71}$$

按 e_1 的分布及 n_k 的定义,当 k 充分大时,有

$$\begin{aligned} P(\widetilde{E}_k) &= P(e_1 \geqslant x_{n_k})^{[2/b_{n_k}]+1} \\ &= (\exp(-C_{n_k}/4))^{[2/b_{n_k}]+1} \\ &\geqslant \exp\left(-\frac{2}{4}(\log k + \log\log k) - C_{n_k}/4\right) \\ &\geqslant \exp\left(-\frac{3}{4}(\log k + \log\log k)\right) \\ &\geqslant \exp(-\log k) = k^{-1}. \end{aligned} \tag{3.72}$$

此处用了由(3.47)得出的 $C_{n_k} \leqslant \log n_k \approx \log k$(当 k 充分大时). 由(3.72),根据 Borel-Cantelli 引理,即得(3.71),因而有(3.70). 这就完成了 $p=1$时定理的证明. 对一般的 p,用 a_i 记由(3.51)所定义的 x_i,然后定义一串 p 维向量 $x_1, x_2, \cdots$ 如下:

$$x_{ip+j} = (0, \cdots 0, a_{i+1}, 0, \cdots, 0)' \quad (1 \leqslant j \leqslant p;\ i = 0, 1, \cdots). \tag{3.73}$$

其中 a_{i+1}在向量 x_{ip+j}中是第 j 个分量. $e_1, e_2, \cdots$的定义如前,则条件(3.42)成立. 但易见 β_0的 LADE $\hat{\beta}_n$ 的任一分量不以概率1收敛于 β_0的相应分量. 定理3.4证毕. ▌

注 由此定理与定理3.2相结合,对 $e_1, e_2, \cdots$满足本定理的条件而言,还遗留下一个情况:

$$d_n = O(1/\log n) \tag{3.74}$$

尚未解决,即:当 $e_1, e_2, \cdots$满足本定理的条件时,(3.74)是否为

LADE 强相合的充分条件?

§3.3 ρ不必为凸函数的情况

本节讨论当 ρ 不必为凸函数时,M 估计的强相合问题.首先要说明一下,为什么在讨论 M 估计的弱相合时,我们只涉及 ρ 为凸函数的情况.这是因为,对不必为凸函数的情况,现今我们知道的方法在大多数情况下(有些不重要的例外)总是或者能证明强相合,或者什么也证明不了.固然强收敛的条件一般总是高于弱收敛,对 ρ 为非凸函数时也应如此,但必须找到其实质差别所在,才有较大的意义.

一、自变量 X 为随机向量的情形

假定有一个 p 维随机向量 X(自变量)和一维随机变量 Y(因变量),把 X 和 Y 联系在一个线性模型内,其意义如何,我们再作一点解释:

一种解释是:对给定的 $X=x$ 考虑 Y 的条件分布 $Y|x$.设此条件分布的某一特征(如期望、中位数之类)为 x 的线性函数 $\alpha_0+x'\beta_0$,作为这样的特征,通常(不必一定)可以找到一个函数 ρ,使得若以 Z_x 记一个其分布与 $Y|x$ 相同的随机变量,则函数 $E\rho(Z_x-\alpha-x'\beta)$(作为 α、β 的函数)在 $\alpha=\alpha_0$和 $\beta=\beta_0$处达到最小.注意,α_0、β_0应与 x 无关.这个解释实际上是一个假定,与 X 非随机的情况相同.

另一种解释是:从线性逼近的角度着眼,有了 X 和 Y,现要用 X 的一个线性函数 $\alpha+X'\beta$ 来逼近 Y,问当 α 和 β 取何值时,逼近最好?而所谓"逼近最好",则理解为对选定的 ρ,作为 α、β 的函数 $E\rho(Y-\alpha-X'\beta)$在 $\alpha=\alpha_0$和 $\beta=\beta_0$处取最小值,则 $\alpha_0+X'\beta_0$逼近最好.

第一种解释是线性模型一词的本意,但如上所说,它实在是一个假定,并非对每个(X,Y)都能这样解释的.第二种解释对(X,Y)没有多少限制,但已不是原来"线性模型"一词的本意了.显然,若

$\alpha_0+X'\beta_0$符合第一种解释，则必符合第二种解释. 因为第一种解释意味着：

$$E(\rho(Y-\alpha_0-X'\beta_0)|X=x)=\min_{\alpha,\beta} E(\rho(Y-\alpha-X'\beta)|X=x), \tag{3.75}$$

对任何 $x\in R^p$，必有

$$E\rho(Y-\alpha_0-X'\beta_0)=\min_{\alpha,\beta} E\rho(Y-\alpha-X'\beta). \tag{3.76}$$

反之，由(3.76)一般不能推出(3.75)对一切的 $x\in R^p$ 成立. 故由 $\alpha_0+X'\beta_0$符合第二种解释一般推不出它符合第一种解释.

由此可见，我们只须在第二种解释的模型下来证明 M 估计的强相合性. 所得结果必然也适用于第一种解释的情形.

下面的定理是本段的主要定理：

定理3.5 设$(X_1,Y_1),(X_2,Y_2),\cdots$是从$(X,Y)$中抽出的 iid. 样本，$(\hat{\alpha}_n,\hat{\beta}_n)$是极值问题

$$\sum_{i=1}^{n}\rho(Y_i-\alpha-x_i'\beta)=\text{最小} \tag{3.77}$$

的一个解(它是作为下文(α_0,β_0)的 M 估计). 假定：

1° ρ 在 R^1 处处连续，且存在 c，使 ρ 在$[c,\infty)$非降，在$(-\infty,c]$非增.

2° $\rho(\infty)=\rho(-\infty)=\infty$.

3° 对任何的 $\alpha\in R^1$及 $\beta\in R^p$，期望

$$Q(\alpha,\beta)=E(\rho(Y-\alpha-X'\beta))$$

存在有限，且 Q 有唯一的最小值点(α_0,β_0).

则当 $n\to\infty$时，有

$$\hat{\alpha}_n\to\alpha_0,\ \hat{\beta}_n\to\beta_0,\ \text{a.s.} \tag{3.78}$$

本定理的证明分解为以下两个引理，不失普遍性，我们在以下总设：

$\alpha_0=0,\beta_0=0$. 又对 $r>0$记 $A_r=[-r,r]^{p+1},\tilde{A}_r=[-r,r]^p$.

引理3.7 设$(\hat{\alpha}_{nl},\hat{\beta}_{nl})$为将$(\alpha,\beta')$局限于 A_l 时极值问题(3.77)的一个解，$l>0$. 则在定理3.5的条件下，有

$$\hat{\alpha}_{nl}^{*} \to 0,\ \hat{\beta}_{nl} \to 0 \text{ a.s. } (\text{当 } n \to \infty). \tag{3.79}$$

证 记 $\rho^*(Y-\alpha-X'\beta)=\rho(Y-\alpha-X'\beta)-\rho(Y)$,

$$Q^*(\alpha,\beta)=E(\rho^*(Y-\alpha-X'\beta)).$$

往下证存在 $K(X,Y)\geqslant 0, E(K(X,Y))<\infty$,且

$$\sup\{|\rho^*(Y-\alpha-X'\beta)|:(\alpha,\beta')\in A_l\}\leqslant K(X,Y). \tag{3.80}$$

事实上,若以 (a_i,b_i') $(1\leqslant i\leqslant T=2^{p+1})$ 记 A_l 的顶点,则由定理的条件1°,易见当 $\rho^*(Y-\alpha-X'\beta)\geqslant 0$ 时,有

$$\rho^*(Y-\alpha-X'\beta)\leqslant \max_{1\leqslant i\leqslant T}|\rho^*(Y-a_i-X'b_i)|.$$

若 $\rho^*(Y-\alpha-X'\beta)<0$,则 $|\rho^*(Y-\alpha-X'\beta)|\leqslant\rho(Y)-\rho(C)$.
取 $K(X,Y)=\max\limits_{1\leqslant i\leqslant T}|\rho^*(Y-a_i-x'b_i)|+\rho(Y)-\rho(C)$ 即得.

根据(3.80)及 ρ 的连续性,由控制收敛定理,知 $Q^*(\alpha,\beta)$ 对 (α,β) 连续. 再根据条件3°(回忆已设 $\alpha_0=0,\beta_0=0$),即知对任何 $\varepsilon\in(0,l)$,有

$$q\equiv\inf\{Q^*(\alpha,\beta):(\alpha,\beta')\in A_l-A_\varepsilon\}>0.$$

取 $\varepsilon_1\in(0,q/5)$,又选充分大的 m,使(I 为指示函数)

$$E(I((X',Y)\overline{\in} A_m)K(X,Y))<\varepsilon_1. \tag{3.81}$$

记 $g=\sup\{|y-a-x'b|:(x',y)\in A_m,(a,b')\in A_l\}$,选取 $\varepsilon_2>0$,使

$$\sup\{|\rho(u_2)-\rho(u_1)|:|u_1|\leqslant g,|u_2|\leqslant g,\ |u_1-u_2|\leqslant\varepsilon_2\}<\varepsilon_1, \tag{3.82}$$

然后选取 $\varepsilon_3>0$,使

$$\sup\{|(a+x'b)-(\tilde{a}+x'\tilde{b})|:(a,b')\in A_l,(\tilde{a},\tilde{b}')\in A_l,\ |a-\tilde{a}|\leqslant\varepsilon_3,\|b-\tilde{b}\|\leqslant\varepsilon_3,\|x\|\leqslant\sqrt{p}\,m\}<\varepsilon_2. \tag{3.83}$$

在 A_l-A_ε 中选一有限集 S,使 A_l-A_ε 中的任一点与 S 的距离不超过 ε_3. 又令

$$J_1(i)=\begin{cases}1,(X_i',Y)\in A_m;\\0,\text{其他},\end{cases}\quad J_2(i)=1-J_1(i), \tag{3.84}$$

则由 q 的定义与(3.81),应用强大数定律,知以概率1当 n 充分大时,对 $(\tilde{\alpha},\tilde{\beta}')\in S$,以下两式同时成立:

$$n^{-1}\sum_{i=1}^{n}\rho^{*}(Y_i-\tilde{\alpha}-x_i'\tilde{\beta})J_1(i)>E\rho^{*}(Y-\tilde{\alpha}-x'\tilde{\beta})$$
$$-E(K(X,Y)I((X',Y)\overline{\in}A_m))-\varepsilon_1\geqslant q-2\varepsilon_1\geqslant 3\varepsilon_1, \quad (3.85)$$

$$n^{-1}\sum_{i=1}^{n}K(X_i,Y_i)J_2(i)<\varepsilon_1. \quad (3.86)$$

由(3.85)和(3.86)得知，对任何$(\alpha,\beta')\in A_l-A_\varepsilon$，有

$$\begin{aligned}
&n^{-1}\sum_{i=1}^{n}\rho^{*}(Y_i-\alpha-X_i'\beta)\\
&\geqslant n^{-1}\sum_{i=1}^{n}\rho^{*}(Y_i-\alpha-X_i'\beta)J_1(i)\\
&\quad-n^{-1}\sum_{i=1}^{n}K(X_i,Y_i)J_2(i)\\
&\geqslant n^{-1}\sum_{i=1}^{n}\rho^{*}(Y_i-\tilde{\alpha}-X_i'\tilde{\beta})J_1(i)\\
&\quad-n^{-1}\sum_{i=1}^{n}\Big|\rho^{*}(Y_i-\alpha-X_i'\beta)\\
&\quad-\rho^{*}(Y_i-\tilde{\alpha}-X_i'\tilde{\beta})\Big|-\varepsilon_1. \quad (3.87)
\end{aligned}$$

此处$(\tilde{\alpha},\tilde{\beta}')\in S_1$满足$|\alpha-\tilde{\alpha}|\leqslant\varepsilon_3$，$\|\beta-\tilde{\beta}\|\leqslant\varepsilon_3$. 按$\varepsilon_2$和$\varepsilon_3$的选法，有

$$|\rho^{*}(Y_i-\tilde{\alpha}-X_i'\tilde{\beta})-\rho^{*}(Y_i-\alpha-X_i'\beta)|$$
$$=|\rho(Y_i-\tilde{\alpha}-X_i'\tilde{\beta})-\rho(Y_i-\alpha-X_i'\beta)|<\varepsilon_1 \quad (1\leqslant i\leqslant n).$$

由此及(3.87)，得

$$n^{-1}\sum_{i=1}^{n}\rho^{*}(Y_i-\alpha-X_i'\beta)$$
$$\geqslant n^{-1}\sum_{i=1}^{n}\rho^{*}(Y_i-\tilde{\alpha}-X_i'\tilde{\beta})J_1(i)-2\varepsilon_1.$$

将此式与(3.85)相结合，即可知以概率1当n充分大时，对一切$(\alpha,\beta')\in A_l-A_\varepsilon$有下式成立：

$$n^{-1}\sum_{i=1}^{n}\rho^{*}(Y_i-\alpha-X_i'\beta)\geqslant\varepsilon_1>0.$$

由此推出：以概率1当n充分大时，有$(\hat{\alpha}_n,\hat{\beta}_n')\in A_\varepsilon$. 由$\varepsilon>0$的任意

性，得(3.79). 引理证毕. ▌

引理3.8 在定理3.5的条件下，设 $\alpha_0=0,\beta_0=0$，则存在$l>0$，使以概率1当 n 充分大时，有$(\hat{\alpha}_n,\hat{\beta}_n)\in A_l$.

证 由条件3°中(α_0,β_0)的唯一性，易见当 α 和 β 不同时为0时，$\alpha+X'\beta$ 不退化为0. 由此可知，若以 S 记 R^{p+1}中的单位球面，则可找到 $\varepsilon>0$，使

$$v=\inf\{P(|\alpha+X'\beta|>\varepsilon):(\alpha,\beta')\in S\}>0. \quad (3.88)$$

取 m 充分大，使 $P(X\in\tilde{A}_m)>1-v/4$. 令 $u=3^{-1}(1+\sqrt{p}\,m)^{-1}\varepsilon$，找 S 的有限子集 S_1，使 S 中的任一点与 S_1的距离不超过 u，由(3.88)，根据强大数定律，知以概率1当 n 充分大时，有

$$\#\{i:1\leqslant i\leqslant n,|\alpha+X_i'\beta|>\varepsilon\}\geqslant nv/2$$
$$(\text{对于一切的}(\alpha,\beta')\in S_1), \quad (3.89)$$

$$\#\{i:1\leqslant i\leqslant n,X_i\in\tilde{A}_m\}\geqslant n(1-v/4) \quad (3.90)$$

(此处$\#\{A\}$记集 A 中所含元素的个数). 找 $L>0$，使 $P(|Y|\leqslant L)>1-v/8$，由强大数定律，知以概率1当 n 充分大时，有

$$\#\{i:1\leqslant i\leqslant n,|Y_i|\leqslant L\}>n(1-v/8). \quad (3.91)$$

取

$$K=8/v, \quad (3.92)$$

由条件1°，知存在 $h>0$，使当 $|a|\geqslant h$，$|u|\leqslant L$ 时，

$$\rho(u+a)-\rho(u)>K. \quad (3.93)$$

现在定义 $l=2h/\varepsilon$，我们来证明：这个 l 是满足本引理中的要求的. 为此，任取 R^{p+1}的一点$(a,b')\overline{\in}A_l$，令$(\alpha,\beta')=(a,b')/r$，其中 $r=\|(a,b')\|>l$，则$(\alpha,\beta')\in S$. 找$(\tilde{\alpha},\tilde{\beta}')\in S_1$，使$|\tilde{\alpha}-\alpha|\leqslant u$，$\|\tilde{\beta}-\beta\|\leqslant u$. 当 $X_i\in A_m$ 且$|\tilde{\alpha}+X_i'\tilde{\beta}|>\varepsilon$时，有

$$\begin{aligned}|\alpha+X_i'\beta|&>\varepsilon-|\alpha-\tilde{\alpha}+X_i'(\beta-\tilde{\beta})|\\&\geqslant\varepsilon-|\alpha-\tilde{\alpha}|-\|X_i\|\|\beta-\tilde{\beta}\|\\&\geqslant\varepsilon-u-m\sqrt{p}\,u=\varepsilon-(1+\sqrt{p}\,m)u>\varepsilon/2.\end{aligned}$$

因而有

$$|a+X_i'b|>r\varepsilon/2>l\varepsilon/2=h.$$

由上式及(3.89)、(3.90)，知以概率1当 n 充分大时，有

$$\#\{i:1\leqslant i\leqslant n,|a+X_i'b|\geqslant h\}\geqslant nv/4\quad(\text{对于一切的}(a,b')\overline{\in}A_l).\tag{3.94}$$

结合(3.91)和(3.94)，知以概率1当 n 充分大时，有

$$\#\{i:1\leqslant i\leqslant n,|a+X_i'b|\geqslant h,|Y_i|\leqslant L\}\geqslant nv/8\quad(\text{对于一切的}(a,b')\overline{\in}A_l).\tag{3.95}$$

由(3.92)、(3.93)和(3.95)，知以概率1当 n 充分大时，有

$$\sum_{i=1}^{n}\rho^*(Y_i-a-X_i'b)\geqslant Knv/8=n\quad(\text{对于一切的}(a,b')\overline{\in}A_l).$$

而这导致了$(\hat{a}_n,\hat{\beta}_n')\in A_l$. 引理证毕. ▌

定理3.5的证明直接由以上两个引理推得. ▌

若 ρ 为凸函数，则定理的条件3°可适当减轻，看以下定理：

定理3.6 设 ρ 为凸函数，期望 $Q^*(\alpha,\beta)=E(\rho^*(Y-\alpha-X'\beta))$对$(\alpha,\beta')\in R^{p+1}$处处存在有限，并且 Q^* 有唯一的最小值点(α_0,β_0')，则(3.78)仍成立.

证 首先易见，由本定理的假定可推出定理3.5的条件1°与2°成立. 据此，从前面的证明可见，只须证明(3.80)($K(X,Y)$与前面的不同)，则以后的论证全部有效.

为证(3.80)，分别以(a_i,b_i') $(1\leqslant i\leqslant T)$和$(a_j,b_j')$ $(T+1\leqslant j\leqslant 2T)$记 A_l 和 A_{2l}的各自 $T=2^{p+1}$个顶点. 往下证：

$$K(X,Y)=2\max_{1\leqslant i\leqslant 2T}|\rho^*(Y-a_i-X'b_i)|$$

满足要求. 首先，$E(K(X,Y))<\infty$不成问题. 现在任取$(x',y)\in R^{p+1}$. 若 $\rho^*(y-\alpha-x'\beta)\geqslant 0$, $(\alpha,\beta')\in A_l$，则由定理3.5的条件1°易见

$$|\rho^*(y-\alpha-x'\beta)|\leqslant\max_{1\leqslant i\leqslant T}|\rho^*(y-a_i-x'b_i)|\leqslant K(X,Y).$$

若 $\rho^*(y-\alpha-x'\beta)<0$，则分两种情况：

$$\alpha+x'\beta>0,\quad\text{或}\quad\alpha+x'\beta<0.$$

两种情况的处理相似，故只取前者即可.

因为$\alpha+x'\beta>0$，由函数 ρ 的凸性，可知 $\rho^*(y-\alpha-x'\beta)$作为

y 的函数是非增的(ρ 为凸函数的假定用在这一关键步骤). 由于$(\alpha,\beta')\in A_t$,知存在 $j\leqslant T$,使 $\alpha+x'\beta\geqslant a_j+x'b_j$. 记 $\tilde{a}=\alpha-a_j$,$\tilde{b}=\beta-b_j$,则 $d\equiv\tilde{a}+x'\tilde{b}\geqslant 0$,故

$$\begin{aligned}\rho^*(y-\alpha-x'\beta)&\geqslant\rho^*(y+d-\alpha-x'\beta)\\&=\rho(y-a_j-x'b_j)-\rho(y+d).\end{aligned}\quad(3.96)$$

由于(α,β')和(a_j,b_j')都属于 A_t,知$(\tilde{a},\tilde{b}')\in A_{2t}$. 由定理3.5的条件1°,知存在 $k\leqslant 2T$,使 $\rho(y+d)\leqslant\rho(y-a_k-x'b_k)$. 以此与(3.96)相结合,并注意到 $\rho^*(y-\alpha-x'\beta)<0$,即得

$$\begin{aligned}|\rho^*(y-\alpha-x'\beta)|&\leqslant|\rho(y-a_j-x'b_j)\\&\quad-\rho(y-a_k-x'b_k)|\\&\leqslant|\rho^*(y-a_j-x'b_j)|\\&\quad+|\rho^*(y-a_k-x'b_k)|\leqslant K(x,y).\end{aligned}$$

这证明了所需要的结果,因而证明了定理3.6. ▌

把本定理用于 LADE,得出以下的结果,由于其重要性,故列为一条定理如下:

定理3.7 设(X_1,Y_1),(X_2,Y_2),…为(X,Y)的 iid. 样本. 假定 $E\|X\|<\infty$,且 $E(|Y-\alpha-X'\beta|-|Y|)$在唯一的一点$(\alpha_0,\beta_0)$达到最小. 则$(\alpha_0,\beta_0)$的 LADE 是强相合的.

在文献[20]中得出过类似本定理的结果,但附加了不必要的条件 $E|Y|<\infty$,$Y-\alpha_0-X'\beta_0$与 X 独立,且当 α 与 β 不同为0时,$P(\alpha+X'\beta=0)=0$.

如果 $\alpha_0+X'\beta_0$符合第一种解释(见§3.3第一段),则我们也可以使用条件化的方法,依据§3.1中的结果,对 ρ 为凸函数而 X 为随机的情况,导出有关(α_0,β_0)的 M 估计的强相合性的一些结果. 这种结果在形式上比定理3.6复杂得多,其条件也比定理3.6的条件苛刻得多,因此没有多大意义. 这一点与前一章所讨论的弱相合不同:对弱相合来说,X 为随机时的定理2.5很容易由 X 为非随机时的定理2.1导出来.

以 LADE 这个例子来说,若 $Y|X=x$ 有中位数 $\alpha_0+X'\beta_0$,则

为要使用定理3.2以导出关于(α_0,β_0)的 LADE 强相合性，就要假定存在与 x 无关的常数 $l_1>0$、$l_2>0$，使对一切 $x\in B$，其中 $P(X\in B)=1$，有

$$P(0<Y-\alpha_0-x'\beta_0<h\,|\,X=x)\geqslant l_1h\quad(0<h<l_2),$$

$$P(-h<Y-\alpha_0-x'\beta_0<0\,|\,X=x)\geqslant l_1h\quad(0<h<l_2),$$

且以概率1有

$$\max_{1\leqslant i\leqslant n}\tilde{x}_i'S_{xn}^{-1}\tilde{x}_i=o(1/\log n)\tag{3.97}$$

成立，其中 $\tilde{x}_i=(1,x_i')'$，$S_{xn}=\sum_{i=1}^{n}\tilde{x}_i\tilde{x}_i'$. 为了保证(3.97)成立，除了 X 的分布不能退化到 R^p 中任一超平面以外，一般须要求：

$$E(\|X\|^2\log\|X\|)<\infty.\tag{3.98}$$

由(3.98)(加上 X 不退化)推出(3.97)的证明留给读者. 可以举出反例证明：即使 $E\|X\|^2<\infty$ 也不能保证(3.97). 一个齐次模型(其中已知 $\alpha_0=0$，不须估计)的反例如下：设 X 为一维，取 $\sqrt{N}$ $(N=1,2,\cdots)$ 为值，当 $N=n^{n^3}$，而 $n=2,3,\cdots$ 时，$P(X=\sqrt{N})=c/(N\sqrt{\log N})$，对其余的 N，则 $P(X=\sqrt{N})=c/N^3$. 选择 $c>0$，使 $\sum_{N=1}^{\infty}P(X=\sqrt{N})=1$. 证明细节也留给读者.

问题　若假定 $\alpha_0+x'\beta_0$ 是 $Y|X=x$ 的唯一中位数，则是否在免去条件 $E\|X\|<\infty$ 时，仍能证明(α_0,β_0)的 LADE 强相合性？有一些理由可信答案是肯定的. 若真如此，那将是有关 X 为随机时 LADE 强相合性最深刻而彻底的结果.

定理3.5处理了 ρ 无界的情况. 当 ρ 有界时，有如下的结果：

定理3.8　保留定理3.5的条件1°和3°，将2°修改为 $\rho(\infty)=\rho(-\infty)<\infty$，且设当 α、β 不同为0时，$P(\alpha+X'\beta=0)=0$. 则(3.78)依旧成立.

证明大体上类似定理3.5，细节上有少许出入，从略.

问题　当 $\rho(\infty)\neq\rho(-\infty)$ 时，情况如何？这一点看起来似非本质的，但迄今尚未得到任何结果.

二、自变量为非随机的情形

在这种情形，没有必要突出常数项 α_0，因此，我们仍回到模型(1.1). 与 X 为随机的情形相比较，在这种情况下所得的结果还很欠完整. 即使在 ρ 为凸函数时，如在§3.1所见，问题离彻底解决也相去甚远. 对 ρ 为非凸函数的情况，当然更是如此.

定理3.9 设在模型(1.1)中，误差 $e_1, e_2, \cdots$ 为 iid.，以 $\hat{\beta}_n$ 记利用函数 ρ 所作的 M 估计. 设满足以下条件：

1° 定理3.5的条件1°、2°成立.

2° 对任何实数 t，$E\rho(e_1+t)<\infty$. 作为 t 的函数，$E\rho(e_1+t)$有唯一最小值点0，且存在常数 $c>0$，使当$|t|$充分小时，有

$$E(\rho(e_1+t)-\rho(e_1))\geqslant ct^2. \tag{3.99}$$

3° $\{X_i\}$有界，且若以 λ_n 记 $S_n=\sum_{i=1}^n x_i x_i'$ 的最小特征根，则

$$\liminf_{n\to\infty} \lambda_n/n>0. \tag{3.100}$$

从而(3.78)成立.

与定理3.5比较，此处条件2°的前一半大体上相当于定理3.5的条件3°. 其后一半，即(3.99)实质是要求0这个唯一最小值点有一定的"鲜明度"，是附加的要求. 即值在弱相合问题中，我们也曾指出过这种附加要求并非是多余的.

重点的部分是条件3°. 从根本上看，自变量为非随机的情况之所以比随机的情况复杂，在于当 X 为随机时，$\{X_i\}$的分布有其概率上的规律性，并非完全杂乱无章. 而在 X 为非随机时，则$\{x_i\}$无这一规律性. 故需要设下一定的限制，以约束其"不良表现". 与定理3.1相比，定理3.9对$\{x_i\}$的条件要强得多. 但另一方面，加在 ρ 上的矩条件则有显著的减弱. 这两方面的条件总是互为消长的.

本定理证明比较复杂，故分解为以下几条引理：

引理3.9 设 e_i 为 iid.，ρ 满足定理3.5的条件1°，且对任何实数 t，$E\rho(e_1+t)<\infty$. 设$\{c_i\}$为一串有界实数，则

$$\lim_{n\to\infty} n^{-1}\sum_{i=1}^n(\rho(e_i+c_i)-E\rho(e_i+c_i))=0,\ \text{a.s.}.$$

证 此引理由 Kolmogorov 强大数定律的下述推广而得出：

设$\{\xi_i\}$为一串独立随机变量，而$\{\eta_i\}$为 iid.，$E|\eta_1|<\infty$，且对一切 i，$|\xi_i|\leqslant|\eta_i|$，则$\{\xi_i\}$服从强大数定律.

　　上述断言的证明如下：给定 $\varepsilon>0$，找充分大的 K，使

$$E(|\eta_1|I(|\eta_1|\geqslant K))\leqslant\varepsilon.$$

将 Kolmogorov 强大数定律用于$\{|\eta_i|I(|\eta_i|\geqslant K),i\geqslant1\}$，有

$$n^{-1}\sum_{i=1}^{n}|\eta_i|I(|\eta_i|\geqslant K)\to E(|\eta_1|I(|\eta_1|\geqslant K))\leqslant\varepsilon,\ \text{a.s.}.$$

另一方面，因为$|\xi_i|\leqslant|\eta_i|$ $(i\geqslant1)$，有

$$|\xi_i|I(|\xi_i|\geqslant K)\leqslant|\eta_i|I(|\eta_i|\geqslant K)\ (i\geqslant1),$$

于是有 $n^{-1}\sum_{i=1}^{n}E(|\xi_i|I(|\xi_i|\geqslant K))\leqslant\varepsilon$，以及

$$\limsup\Big(n^{-1}\sum_{i=1}^{n}|\xi_i|I(|\xi_i|\geqslant K)\Big)\leqslant\varepsilon,\ \text{a.s.}.$$

因为随机变量序列$\{\xi_iI(|\xi_i|<K),i\geqslant1\}$一致有界，故必满足强大数定律. 因此

$$J_n\equiv n^{-1}\Big(\sum_{i=1}^{n}\xi_iI(|\xi_i|<K)-\sum_{i=1}^{n}E(\xi_iI(|\xi_i|<K))\Big)\to0,\ \text{a.s.}.$$

总结以上诸式，并注意

$$\begin{aligned}
&n^{-1}\Big|\sum_{i=1}^{n}\xi_i-\sum_{i=1}^{n}E\xi_i\Big|\\
&\quad\leqslant n^{-1}\Big|\sum_{i=1}^{n}\xi_iI(|\xi_i|<K)-\sum_{i=1}^{n}E(\xi_iI(|\xi_i|<K))\Big|\\
&\qquad+n^{-1}\sum_{i=1}^{n}E(|\xi_i|I(|\xi_i|\geqslant K))\\
&\qquad+n^{-1}\sum_{i=1}^{n}|\xi_i|I(|\xi_i|\geqslant K),
\end{aligned}$$

得

$$\limsup\Big(n^{-1}\Big|\sum_{i=1}^{n}\xi_i-\sum_{i=1}^{n}E\xi_i\Big|\Big)\leqslant\limsup J_n+\varepsilon+\varepsilon=2\varepsilon,\ \text{a.s.}.$$

因为这一结论对任给的 $\varepsilon>0$都成立，从而证明了上述断言.

　　现在不难证明本引理. 记 $c_o=\sup\limits_{n\geqslant1}|c_n|$. 因为函数 ρ 满足定理

3.5的条件1°,故有

$$\rho(e_i + c_i) - \rho(c) \leqslant (\rho(e_i + c_0) - \rho(c)) + (\rho(e_i - c_0) - \rho(c)).$$

注意到 c 是 ρ 的最小值点,可知若把上式左、右两边分别记为 ξ_i 和 η_i,则以上证明的断言可用. 从而得到所要的结论. ▌

引理3.10 设函数 ρ 满足定理3.9的条件1°和2°,$\{x_i\}$为有界点列,B 为有界子集,e_i 为 iid., 则下述断言:"函数列

$$n^{-1}\sum_{i=1}^{n}\left(\rho(e_i - x_i'\beta) - E\rho(e_i - x_i'\beta)\right) \quad (n \geqslant 1)$$

在 B 上为等度连续"以概率 1 成立.

证 以 F 记 e_1 的分布,作概率空间$(R^\infty, \mathscr{B}^\infty, F^\infty)$,其中 $\mathscr{B}^\infty$ 是 R^∞ 中一切 Borel 集所构成的 σ-域. 记

$$T = \sup\{|x_i'\beta| : i \geqslant 1, \ \beta \in B\}.$$

根据假定,有 $T < \infty$. 现在固定自然数 m,而找常数 $h > 0$,使

$$E(\rho(e_1 + T)I(\rho(e_1 + T) \geqslant h)) + E(\rho(e_1 - T)I(\rho(e_1 - T) \geqslant h)) < (3m)^{-1}. \quad (3.101)$$

根据对 ρ 的假定,可找到 $\varepsilon_m' > 0$,使

$$|\rho(u_1) - \rho(u_2)| \leqslant (3m)^{-1}, \text{当} |u_1 - u_2| \leqslant \varepsilon_m' \text{ 且 } \min(\rho(u_1), \rho(u_2)) \leqslant h \text{ 时}. \quad (3.102)$$

再找 $\varepsilon_{m1} > 0$,使当 β、$\tilde{\beta}$ 都属于 B,而距离不超过 ε_{m1}时,有

$$\sup_i |x_i'(\beta - \tilde{\beta})| \leqslant \varepsilon_m'.$$

由(3.101),使用强大数定律,可找到自然数 N_m 及集 $D_m \in \mathscr{B}^\infty$,满足 $F^\infty(D_m) < 2^{-m}$,使当 $n \geqslant N_m$ 及 $e^* = (e_1, e_2, \cdots) \overline{\in} D_m$时,有

$$n^{-1}\sum_{i=1}^{n}\rho(e_i + T)I(\rho(e_i + T) \geqslant h) + n^{-1}\sum_{i=1}^{n}\rho(e_i - T)I(\rho(e_i - T) \geqslant h) < (3m)^{-1}. \quad (3.103)$$

因为$\{x_i\}$和 B 都有界而 ρ 连续,对任意的 $e^+ \in R^\infty$,可找到 $\varepsilon_{m2}(e^*) > 0$,使当 β、$\tilde{\beta}$ 都属于 B 而其距离不超过 $\varepsilon_{m2}(e^*)$时,对$1 \leqslant n \leqslant N_m$

同时有下式成立：

$$\left|n^{-1}\sum_{i=1}^{n}(\rho(e_i-x_i'\bar{\beta})-\rho(e_i-x_i'\beta))\right|<(3m)^{-1}. \tag{3.104}$$

取
$$\varepsilon_m(e^*)=\min(\varepsilon_{m1},\varepsilon_{m2}(e^*)),$$
现设 $e^*\overline{\in}D_m$，β 与 $\bar{\beta}$ 都属于 B，且距离不超过 $\varepsilon_m(e^*)$，则当 $1\leqslant n\leqslant N_m$时，有(3.104). 若 $n>N_m$，则注意到 ρ 为非负，有

$$\left|n^{-1}\sum_{i=1}^{n}(\rho(e_i-x_i'\bar{\beta})-\rho(e_i-x_i'\beta))\right|\leqslant J_1+J_2+J_3, \tag{3.105}$$

其中

$$J_1=n^{-1}\sum_{i=1}^{n}\rho(e_i-x_i'\bar{\beta})I(\rho(e_i-x_i'\bar{\beta})\geqslant h);$$

$$J_2=n^{-1}\sum_{i=1}^{n}\rho(e_i-x_i'\beta)I(\rho(e_i-x_i'\beta)\geqslant h);$$

$$J_3=n^{-1}\left|\sum{}'(\rho(e_i-x_i'\bar{\beta})-\rho(e_i-x_i'\beta))\right|.$$

J_3中的 $\sum'$ 表示求和的范围为：$1\leqslant i\leqslant n$，且
$$\min(\rho(e_i-x_i'\beta),\rho(e_i-x_i'\bar{\beta}))<h.$$
按前面的论证，J_1和 J_2都不超过(3.103)的左端，故都不超过$(3m)^{-1}$. 而根据 ε_m'、ε_{m1}及 $\varepsilon_m(e^*)$的定义以及(3.102)，J_3的和中每一项(不计因子 n^{-1})都不超过$(3m)^{-1}$，故 $J_3\leqslant(3m)^{-1}$. 综合上述可知，对每个 $e^*\overline{\in}D_m$，存在 $\varepsilon_m(e^*)$，使当 β 和 $\bar{\beta}$ 都属于 B 且其距离不超过 $\varepsilon_m(e^*)$时，对任何 $n\geqslant1$(3.104)成立，但右端的$(3m)^{-1}$要改为 m^{-1}.

现令 $D=\bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty}D_m$，则 $F^{\infty}(D)=0$，且当 $e^*\overline{\in}D$ 时，对充分大的 m，有 $e^*\overline{\in}D_m$. 由此可知，当 $e^*\overline{\in}D$ 时，函数列

$$\left\{\sum_{i=1}^{n}\rho(e_i-x_i'\beta)/n:n\geqslant1\right\}$$

等度连续，即以概率1为等度连续. 故为证本引理，只须证明函数列

$\left\{\sum_{i=1}^{n} E\rho(e_i - x_i'\beta)/n : n \geqslant 1\right\}$ 在 B 上等度连续. 这立即由 e_i 为 iid.,$E\rho(e_1+c)$作为 c 的函数在 R^1处处连续,以及 B 和$\{x_i\}$都有界可推出来. 引理3.10证毕. ▌

结合引理3.9和3.10,经过一些例行推导,即可得到以下引理:

引理3.11 设引理3.9和3.10的条件都满足,则以概率1(当 $e^{\cdot}\overline{\in}D$)成立:当 $n\to\infty$时, $\sum_{i=1}^{n}(\rho(e_i-x_i'\beta)-E\rho(e_i-x_i'\beta))/n$ 在 $\beta\in B$ 上一致收敛于0.

引理3.12 设定理3.9的条件2°、3°满足,而 B 为 R^p 的有界子集,$\varepsilon>0$. 则

1° 存在只与 B 有关的常数 $b>0$,使当 n 充分大时,有

$$\sum_{i=1}^{n} E(\rho(e_i - x_i'\beta) - \rho(e_i))/n \geqslant b\|\beta\|^2 \quad (\text{当 } \beta\in B);$$

2° 对充分小的 $\varepsilon'>0$,$\liminf_{n\to\infty}\{\inf\{\#(i:1\leqslant i\leqslant n, |x_i'\beta|>\varepsilon'):\|\beta\|\geqslant\varepsilon\}/n\}>0$.

证 1° 根据定理3.9的条件2°以及$\{x_i\}$和 B 皆有界,易见存在常数 $c>0$,使 $E(\rho(e_i-x_i'\beta)-\rho(e_i))\geqslant c(x_i'\beta)^2$. 于是

$$\sum_{i=1}^{n} E(\rho(e_i - x_i'\beta) - \rho(e_i))/n \geqslant c\beta'\left(\sum_{i=1}^{n} x_i x_i'/n\right)\beta. \quad (3.106)$$

按(3.100),知存在 $\lambda>0$,使当 n 充分大时,$\sum_{i=1}^{n} x_i x_i' \geqslant \lambda n I_p$. 故由(3.106)得知当 n 充分大时,

$$\sum_{i=1}^{n} E(\rho(e_i - x_i'\beta) - \rho(e_i))/n \geqslant c\lambda\|\beta\|^2 = b\|\beta\|^2$$

对 $\beta\in B$ 同时成立.

2° 用反证法:若结论不成立,则存在自然数子列$\{n'\}$,R^p 中点列$\{\beta_{n'}\}$满足$\|\beta_{n'}\|\geqslant\varepsilon$,及一串 $\varepsilon_{n'}\downarrow 0$,使当 $n'\to\infty$时,有

$$c_{n'} \equiv \#(i:1\leqslant i\leqslant n_i;\ |x_i'\beta_{n'}| > \varepsilon_{n'})/n' \to 0.$$

记 $M=\sup\{\|x_i\|^2, i\geqslant 1\}$,有

$$\sum_{i=1}^{n'}\beta_{n'}'x_ix_i'\beta_{n'}\leqslant n'\varepsilon_{n'}^2+n'c_{n'}M\|\beta_{n'}\|^2.$$

从而有

$$\lambda_{n'}\|\beta_{n'}\|^2\leqslant n'\varepsilon_{n'}^2+n'c_{n'}M\cdot\|\beta_{n'}\|^2,$$

注意到$\|\beta_{n'}\|\geqslant\varepsilon$，$\varepsilon_{n'}\downarrow 0$及$c_{n'}\to 0$，则上式与(3.100)矛盾，从而有结论2°. 引理证毕. ▍

现在不难完成定理3.9的证明了. 证明的思路仍是与定理3.5一样，所证引理3.7和3.8在此处仍有效. 但我们把那里的集合A_l修改为R^p中以0点为中心，l为半径的球体$B_l=\{\beta:\|\beta\|\leqslant l\}$.

任取$\varepsilon\in(0,l)$，因为

$$\begin{aligned}\sum_{i=1}^{n}(\rho(e_i-x_i'\beta)-\rho(e_i))/n=&\sum_{i=1}^{n}(\rho(e_i-x_i'\beta)-E\rho(e_i-x_i'\beta))/n\\&-\sum_{i=1}^{n}(\rho(e_i)-E\rho(e_i))/n\\&+\sum_{i=1}^{n}(E\rho(e_i-x_i'\beta)-E\rho(e_i))/n\\=&J_1+J_2+J_3,\end{aligned}$$

根据引理3.9，知以概率1成立：当$n\to\infty$时，J_1和J_2对$\beta\in A_l-A_\varepsilon$一致地收敛于0. 而根据引理3.12的1°，当$n\to\infty$时，对$\beta\in A_l-A_\varepsilon$一致地有$J_3\geqslant c$ (c为某个大于0的常数). 由此可知，以概率1成立：当n充分大时，有

$$\inf\left\{\sum_{i=1}^{n}(\rho(e_i-x_i'\beta)-\rho(e_i)):\beta\in A_l-A_\varepsilon\right\}>0,$$

而此式意味着当n充分大时，$\|\hat{\beta}_{nl}\|<\varepsilon$，$\hat{\beta}_{nl}$与引理3.7中的$\hat{\beta}_{nl}$相同(只是$A_l$改为现在的$B_l$). 由$\varepsilon>0$的任意性，可知引理3.7的结论适用于此处的情况.

引理3.8也适用于此处，其证明与前而X为随机时的情况相同. 事实上，细察引理3.8的证明，可见关键在于定理3.5的条件3°中(α_0,β_0)的唯一性及(3.88)式这两点. 前一点的作用由定理3.9的条件2°起到，而后一点的作用则由引理3.12的2°起到，细节上完

全相似,在此不赘述了.定理3.9证毕. ▎

注 如果模型带常数项 $Y_i=\alpha_0+x_i'\beta_0+e_i$,则除了可直接应用定理3.9以外,还可用如下的修改形式:记 $\bar{x}_n=\sum_{i=1}^{n}x_i/n$,以 $\tilde{\lambda}_n$ 记 $\sum_{i=1}^{n}(x_i-\bar{x}_n)(x_i-\bar{x}_n)'$ 的最小特征根,而把定理中的条件3°改为

$$\liminf_{n\to\infty}\ \tilde{\lambda}_n/n>0, \tag{3.107}$$

证明唯一需要改动之处是引理3.12的2°要用

$$\liminf_{n\to\infty}\ \{\inf(\#(i:1\leqslant i\leqslant n,|\alpha+x_i'\beta|>\varepsilon'):\|(\alpha,\beta')\|\geqslant\varepsilon)/n\}>0$$

来代替.这是很容易由(3.107)得到证明的.

第 4 章

M 估计的渐近正态性

考虑线性回归模型(1.1),即模型

$$Y_i = x_i'\beta_0 + e_i \quad (i = 1,\cdots,n), \tag{4.1}$$

其中 Y_i 和 e_i 分别为第 i 次观测值和试验误差,$\{x_i\}$为已知的 p 维点列,β_0 是未知的 p 维回归参数向量.

如前所述,为估计 β_0,可由(1.5)式定义它的一个 M 估计 $\hat{\beta}_n$,即对选定的函数 $\rho: R^1 \to R^1$,定义 $\hat{\beta}_n$ 为极值问题

$$\sum_{i=1}^{n} \rho(Y_i - x_i'\hat{\beta}_n) = \inf_{\beta \in R^p} \sum_{i=1}^{n} \rho(Y_i - x_i'\beta) \tag{4.2}$$

的一个解.或者,也可以像某些文献中那样,由估计方程(1.30),即

$$\sum_{i=1}^{n} \Psi(Y_i - x_i'\hat{\beta}_n)x_i = 0 \tag{4.3}$$

给出 β_0 的 M 估计 $\hat{\beta}_n$ 的另一种定义,此处 $\Psi: R^1 \to R^1$ 为一个选定的函数.本章仅限于讨论 ρ 为(非单调的)凸函数及 Ψ 为增函数(不为常数)的情形.正如§1.2 中所指出的,在 Ψ 有间断的情形,方程(4.3)不一定有解.因此,代替(4.3)式,本章中将只要求 $\hat{\beta}_n$ 满足以下较弱的条件:

$$\left\| S_n^{-1/2} \sum_{i=1}^{n} \Psi(Y_i - x_i'\hat{\beta}_n)x_i \right\| = o_p(1) \quad (当\ n \to \infty) \tag{4.4}$$

此处我们已假定当 n 充分大时,

$$S_n = \sum_{i=1}^{n} x_i x_i'$$

为 p 阶正定方阵.

在前三章中,我们研究了 β_0 的 M 估计的强弱相合性.在本章中,我们要研究它们的渐近正态性,下一章将研究 M 检验统计量的极限分布.当我们用 M 方法对 β_0 作统计推断时,这两章将为有关的大样本推断提供一个理论依据.当然,为了作出这样的统计推断,还需要寻求多余参数的相合估计,这将留待第5章讨论.我们将在§4.1和§4.2中分别建立两种 M 估计的渐近正态性,在§4.3中介绍有关的历史背景.由于第4、5两章需要较多的凸函数的知识,我们将在附录中作一系统的介绍.

我们总假定 $\hat{\beta}_n$ 是可测的,即它们都是随机变量.

§4.1 由极值定义的 M 估计的情形

在本节中,我们只研究由(4.2)式定义的 M 估计 $\hat{\beta}_n$ 的渐近正态性.为此,引进下述假定:

A4.1 假定 $e_1,e_2,\cdots$ 独立同分布(iid.),且 e_1 的分布为 F.

A4.2 设 $\rho: R^1\to R^1$ 为一凸函数,Ψ_-、Ψ_+ 分别为它的左、右导函数,Ψ 为介于 Ψ_- 和 Ψ_+ 之间的一个函数,即

$$\Psi_-(u)\leqslant\Psi(u)\leqslant\Psi_+(u)\quad(\text{对任意的 } u\in R).$$

A4.3 函数 $G(u)\equiv E\Psi(e_1+u)$ 在 $u=0$ 附近存在、有限,$E\Psi(e_1)=G(0)=0$,且 G 在 $u=0$ 有正的导数 λ.

A4.4 $0<E\Psi^2(e_1)=\sigma^2<\infty$,且

$$\lim_{u\to 0}E[\Psi(e_1+u)-\Psi(e_1)]^2=0.$$

A4.5 存在 n_0,当 $n\geqslant n_0$ 时,

$$S_n=\sum_{i=1}^n x_ix_i'$$

为 p 阶正定方阵,且

$$d_n\equiv\max_{1\leqslant i\leqslant n}x_i'S_n^{-1}x_i\to 0\quad(\text{当 } n\to\infty).$$

以下我们总假定 $n\geqslant n_0$.我们有下述重要定理:

定理 4.1 设在模型(4.1)中,A4.1~A4.5成立.则

$$S_n^{1/2}(\hat{\beta}_n - \beta_0) \xrightarrow{\mathscr{L}} N(0, \lambda^{-2}\sigma^2 I_p),$$

此处，I_p 为 p 阶单位方阵，$\xrightarrow{\mathscr{L}}$表示依分布收敛.

关于这个定理的证明，我们将采用 Bai、Rao 和 Wu 在文献[12]中及 Rao 和 Zhao 在文献[55]中所用的简单证明方法. 为此，令

$$\beta_0(n) = S_n^{1/2}\beta_0,\ x_{ni} = S_n^{-1/2}x_i \quad (i = 1, \cdots, n),$$

并将模型(4.1)写成下述形式：

$$Y_i = x_{ni}'\beta_0(n) + e_i \quad (i = 1, \cdots, n), \tag{4.5}$$

其中 x_{ni}满足条件

$$\sum_{i=1}^{n} x_{ni}x_{ni}' = I_p \tag{4.6}$$

及

$$d_n \equiv \max_{1 \leqslant i \leqslant n} \| x_{ni} \|^2 \to 0 \quad (\text{当 } n \to \infty). \tag{4.7}$$

易见，$\hat{\beta}(n) \equiv S_n^{1/2}\hat{\beta}_n$ 即为模型(4.5)之下 $\beta_0(n)$的一个 M 估计.

这里需要下述引理：

引理 4.1　设 D 为 R^p 的一个开凸子集，$\{f_n\}$是 D 上的一列随机凸函数，当 $n \to \infty$时，对每个 $u \in D$，

$$f_n(u) \to f(u),\quad \text{a.s.(或依概率)}.$$

此处 f 是 D 上的一个实函数，则 f 在 D 上也是凸的，且对 D 的每个紧子集 D_0，当 $n \to \infty$时，

$$\sup_{u \in D_0} |f_n(u) - f(u)| \to 0,\quad \text{a.s.(或依概率)}.$$

又若 f 在 D 上是可微的，$\zeta(u)$和 $\zeta_n(u)$分别表示 f 在 u 的梯度和 f_n 在 u 的一个次梯度，则上述收敛蕴涵了对每个这样的 D_0，有

$$\sup_{u \in D_0} \| \zeta_n(u) - \zeta(u) \| \to 0,\quad \text{a.s.(或依概率)}.$$

证　引理 4.1 的 a.s. 版本是附录中凸函数的性质 3)和 9)的直接推论. 今对依概率收敛的情形按 T. Brown 指出的"对角线法"加以证明(参看 Andersen 和 Gill 的文献[8]). f 的凸性是显然的. 今设 $\{u_1, u_2, \cdots\}$为 D 的一个可数稠密子集. 因为 $f_n(u_1) \xrightarrow{P.}$

$f(u_1)$，故对$\{f_n\}$的每个子列，可从中找出一个子列$\{f_{1,k}\}$，使得

$$\lim_{k\to\infty} f_{1,k}(u_1) = f(u_1)\quad \text{a. s.}.$$

同样，存在$\{f_{1,k}\}$的一个子列$\{f_{2,k}\}$，使

$$\lim_{k\to\infty} f_{2,k}(u_2) = f(u_2)\quad \text{a. s.}.$$

一般地，存在$\{f_{m-1,k}\}$的一个子列$\{f_{m,k}\}$，使得

$$\lim_{k\to\infty} f_{m,k}(u_m) = f(u_m),\quad \text{a. s.}.$$

今按"对角线法"作子列$\{f_{m,m}\}$，则

$$\lim_{m\to\infty} f_{m,m}(u_i) = f(u_i),\ \text{a. s.}\quad (i = 1,2,\cdots).$$

由附录中凸函数的性质 3)和 9)，对 D 的每个紧子集 D_0，当 $m\to\infty$ 时，

$$\sup_{u\in D_0} |f_{m,m}(u) - f(u)| \to 0,\quad \text{a. s.} \tag{4.8}$$

且

$$\sup_{u\in D_0} \| \zeta_{m,m}(u) - \zeta(u) \| \to 0,\quad \text{a. s.}. \tag{4.9}$$

这样就证明了，对于任何子列，我们可从中选出一个子列$\{f_{m,m}\}$和$\{\zeta_{m,m}\}$，使得(4.8)和(4.9)式都成立，由依概率收敛的性质，当 $n\to\infty$时，

$$\sup_{u\in D_0} |f_n(u) - f(u)| \xrightarrow{P.} 0,$$

且

$$\sup_{u\in D_0} \| \zeta_n(u) - \zeta(u) \| \xrightarrow{P.} 0.$$

这就完成了引理的证明.　▌

引理 4.2　设在模型(4.1)中，A4.1～A4.5 成立，则对任何常数 $C>0$，

$$\sup_{|\gamma|\leqslant C} \left| \sum_{i=1}^{n} \{\rho(e_i - x_{ni}'\gamma) - \rho(e_i) + \Psi(e_i)x_{ni}'\gamma\} - \frac{1}{2}\lambda\gamma'\gamma \right| \xrightarrow{P.} 0.$$

$$\sup_{|\gamma|\leqslant C} \left| \sum_{i=1}^{n} \{\Psi(e_i - x_{ni}'\gamma) - \Psi(e_i)\} x_{ni} + \lambda\gamma \right| \xrightarrow{P.} 0.$$

正如第 1 章开头指出的，$|\gamma|$表示向量 γ 的各分量的绝对值的

最大值.

证　记

$$f_n(\gamma)=\sum_{i=1}^{n}\{\rho(e_i-x_{ni}'\gamma)-\rho(e_i)+\Psi(e_i)x_{ni}'\gamma\}$$

$$=\sum_{i=1}^{n}\int_0^{-x_{ni}'\gamma}\{\Psi(e_i+u)-\Psi(e_i)\}\mathrm{d}u\quad(\gamma\in R^p).$$

对每个固定的 γ，由 $d_n\to 0$，知 $\max\limits_{1\leqslant i\leqslant n}|x_{ni}'\gamma|\to 0$. 再由假定 A4.3 及规范化条件(4.6)，存在一串正数 $\varepsilon_n\to 0$ 及 $\theta_{ni}\in(-1,1)$，使

$$Ef_n(\gamma)=\sum_{i=1}^{n}\int_0^{-x_{ni}'\gamma}G(u)\mathrm{d}u$$

$$=\sum_{i=1}^{n}\int_0^{-x_{ni}'\gamma}\{\lambda u+o(|u|)\}\mathrm{d}u$$

$$=\frac{1}{2}\lambda\sum_{i=1}^{n}(x_{ni}'\gamma)^2(1+\varepsilon_n\theta_{ni})\to\frac{1}{2}\lambda\gamma'\gamma.$$

由 Schwarz 不等式、A4.4 及 $\max\limits_{1\leqslant i\leqslant n}|x_{ni}'\gamma|\to 0$，

$$\mathrm{Var}\{f_n(\gamma)\}\leqslant\sum_{i=1}^{n}E\left\{\int_0^{-x_{ni}'\gamma}[\Psi(e_i+u)-\Psi(e_i)]\mathrm{d}u\right\}^2$$

$$\leqslant\sum_{i=1}^{n}|x_{ni}'\gamma|\cdot\left|\int_0^{-x_{ni}'\gamma}E[\Psi(e_i+u)-\Psi(e_i)]^2\mathrm{d}u\right|$$

$$=o(1)\cdot\sum_{i=1}^{n}(x_{ni}'\gamma)^2\to 0.$$

因此，对每个 $\gamma\in R^p$，

$$f_n(\gamma)\xrightarrow{P.}\frac{\lambda}{2}\gamma'\gamma,\text{当 } n\to\infty.$$

利用引理 4.1 即得引理 4.2. 证毕.　▌

下面我们来证明定理 4.1：

证　不失一般性，可设模型(4.1)中的真参数 $\beta_0=0$，亦即模型(4.5)中的 $\beta_0(n)=0$. 令

$$\bar{\beta}(n)=\lambda^{-1}\sum_{i=1}^{n}\Psi(e_i)x_{ni}.$$

容易看出,在定理 4.1 的条件下,Lindeberg 条件成立,因而

$$\overline{\beta}(n) \xrightarrow{\mathscr{L}} N(0, \lambda^{-2}\sigma^2 I_p), \text{当 } n \to \infty. \tag{4.10}$$

由此得出,$\{\overline{\beta}(n)\}$是依概率有界的,即对$\forall\ \varepsilon>0$,存在常数 $C>0$,当 $n \geqslant n_0$ 时,

$$P\{\ \overline{\beta}(n)| > C\} < \varepsilon/2. \tag{4.11}$$

由引理 4.2,对任给的 $\delta>0$,

$$I(|\overline{\beta}(n)| \leqslant C) \sup_{\|\gamma-\overline{\beta}(n)\|=\delta} \left| \sum_{i=1}^{n} \{\rho(Y_i - x'_{ni}\gamma) - \rho(Y_i - x'_{ni}\overline{\beta}(n))\} - \frac{1}{2}\lambda \|\gamma - \overline{\beta}(n)\|^2 \right| \xrightarrow{P.} 0, \text{当 } n \to \infty. \tag{4.12}$$

此处 $I(A)$表示集合 A 的示性函数.由(4.11)及(4.12)式,当 n 充分大时,

$$P\left\{ \inf_{\|\gamma-\overline{\beta}(n)\|=\delta} \sum_{i=1}^{n} \rho(Y_i - x'_{ni}\gamma) \geqslant \sum_{i=1}^{n} \rho(Y_i - x'_{ni}\overline{\beta}(n)) + \lambda\delta^2/4 \right\} \geqslant 1 - \varepsilon.$$

由此及 ρ 的凸性,当 n 充分大时,

$$P\{\ \|\hat{\beta}(n) - \overline{\beta}(n)\| < \delta\} \geqslant 1 - \varepsilon.$$

因为 ε 和 δ 都是任意的,故

$$\hat{\beta}(n) - \overline{\beta}(n) \xrightarrow{P.} 0, \text{当 } n \to \infty. \tag{4.13}$$

定理 4.1 即由(4.10)及(4.13)式得到.证毕. ▌

在实际应用中,回归模型中常常带有常数项,因此我们需要考虑以下的线性回归模型:

$$Y_i = \alpha_0 + x'_i\beta_0 + e_i \quad (i = 1, \cdots, n). \tag{4.14}$$

虽然(4.14)为(4.1)的特例,可用定理 4.1 处理,但为了统计推断的目的,还是写成下面形式的定理更为方便:

定理 4.2 设在模型(4.14)中,$(\hat{\alpha}_n, \hat{\beta}_n')$为$(\alpha_0, \beta_0')$的 M 估计.记

$$\bar{x}_n = n^{-1}\sum_{i=1}^{n} x_i, \quad T_n = \sum_{i=1}^{n}(x_i - \bar{x}_n)(x_i - \bar{x}_n)'.$$

设 A4.1～A4.4 成立，且有

A4.5' 存在 n_0，当 $n \geqslant n_0$ 时，T_n 为 p 阶正定方阵，且

$$\max_{1\leqslant i\leqslant n}(x_i - \bar{x}_n)' T_n^{-1}(x_i - \bar{x}_n) \to 0，当 n \to \infty，$$

则

$$\lambda\sigma^{-1} T_n^{1/2}(\hat{\beta}_n - \beta_0) \xrightarrow{\mathscr{L}} N(0, I_p), \tag{4.15}$$

$$\lambda\sigma^{-1}\sqrt{n}(1 + n\bar{x}_n' T_n^{-1}\bar{x}_n)^{-1/2}(\hat{\alpha}_n - \alpha_0) \xrightarrow{\mathscr{L}} N(0,1), \tag{4.16}$$

且当 $n\to\infty$ 时，$T_n^{1/2}(\hat{\beta}_n - \beta_0)$ 与 $\sqrt{n}[(\hat{\alpha}_n - \alpha_0) + \bar{x}_n'(\hat{\beta}_n - \beta_0)]$ 是渐近独立的.

证 定义

$$\beta_0(n) = T_n^{1/2}\beta_0, \ \alpha_0(n) = \sqrt{n}(\alpha_0 + \bar{x}_n'\beta_0),$$
$$\gamma_0(n) = (\alpha_0(n), \beta_0'(n))',$$
$$x_{ni} = T_n^{-1/2}(x_i - \bar{x}_n),$$
$$z_{ni} = (n^{-1/2}, x_{ni}')' \quad (i = 1, \cdots, n).$$

模型(4.14)可转化为

$$Y_i = z_{ni}'\gamma_0(n) + e_i \quad (i = 1, \cdots, n). \tag{4.17}$$

由于 $\sum_{i=1}^{n} z_{ni}z_{ni}' = I_{p+1}$，且 A4.5' 保证了 $\max_{1\leqslant i\leqslant n}\|z_{ni}\| \to 0$，定理 4.1 可用. 设 $\hat{\gamma}(n) = (\hat{\alpha}(n), \hat{\beta}(n)')'$ 为模型(4.17)中 $\gamma_0(n)$ 的 M 估计，则

$$\lambda\sigma^{-1}(\hat{\gamma}(n) - \gamma_0(n)) \xrightarrow{\mathscr{L}} N(0, I_{p+1})，当 n\to\infty.$$

由此知(4.15)式成立，且 $\hat{\alpha}(n) = \sqrt{n}(\hat{\alpha}_n + \bar{x}_n'\hat{\beta}_n)$ 与 $\hat{\beta}(n) = T_n^{1/2}\hat{\beta}_n$ 渐近独立. 由此不难得出(4.16)式. 证毕. ▌

下面我们考虑 M 估计的一种重要的特殊情形，即最小绝对偏差估计(LADE)或称最小一乘估计，它在 §1.1 的第二段中已有定义.

设在模型(4.1)中，A4.1 成立，且 0 为 F 的中位数. 此时为了估计 β_0，可采用它的最小一乘估计 $\hat{\beta}_n$，它定义为下述极值问题的

一个解：

$$\sum_{i=1}^{n}|Y_i - x_i'\hat{\beta}_n| = \inf_{\beta\in R^p}\sum_{i=1}^{n}|Y_i - x_i'\beta|. \tag{4.18}$$

关于 β_0 的最小一乘估计 $\hat{\beta}_n$ 的渐近正态性，我们有下述定理，它可由定理 4.1 直接得出.

定理 4.3 设在模型(4.1)之下，$e_1, e_2, \cdots$ 独立同分布，其共同分布为 F，中位数为 0，F 在 $u=0$ 有正的导数 $f(0)$，且 A4.5 成立. 则

$$2f(0)S_n^{1/2}(\hat{\beta}_n - \beta_0) \xrightarrow{\mathscr{L}} N(0, I_p)，当\ n\to\infty.$$

在模型有常数项时，关于最小一乘估计的渐近正态性，也有与定理 4.2 类似的下述结论：

定理 4.4 设在模型(4.14)中，$e_1, e_2, \cdots$ 独立同分布，其共同分布为 F，中位数为 0，F 在零点有正的导数 $f(0)$，且 A4.5* 成立. 设 $(\hat{\alpha}_n, \hat{\beta}_n)$ 为 (α_0, β_0) 的最小一乘估计，则

$$2f(0)T_n^{1/2}(\hat{\beta}_n - \beta_0) \xrightarrow{\mathscr{L}} N(0, I_p),$$

$$2f(0)\sqrt{n}(1 + n\bar{x}_n'T_n^{-1}\bar{x}_n)^{-1/2}(\hat{\alpha}_n - \alpha_0) \xrightarrow{\mathscr{L}} N(0,1),$$

且当 $n\to\infty$ 时，$T_n^{1/2}(\hat{\beta}_n - \beta_0)$ 与 $\sqrt{n}[(\hat{\alpha}_n - \alpha_0) + \bar{x}_n'(\hat{\beta}_n - \beta_0)]$ 是渐近独立的.

定理 4.3 和 4.4 是陈希孺、白志东、赵林城和吴月华于 1990 年在文献[4]中首次严格地证明了的. 他们用的是另一种证明方法. 后来，Pollard 于 1991 年在文献[50]中也得到了相同的结果.

下面我们仔细地讨论一下定理 4.1 的条件. 正如同(3.11)式所指出的，$d_n\to 0$ 蕴涵了 $S_n^{-1}\to 0$，反之不真. 因此，若将定理 4.1 与定理 2.1 的条件加以比较，不难看出我们为得到 $\hat{\beta}_n$ 的渐近正态性所加的条件比相合性的条件强一些.

另一方面，正如 Huber(1973，文献[34]中)所指出的，在模型(4.1)中，对 β_0 的最小二乘估计(LSE)的渐近正态性而言，在某些附加条件下，条件 A4.5(即 $d_n\to 0$)不仅是充分的，而且也是必要的. 换言之，我们有如下定理：

定理 4.5　假定在模型(4.1)中，$e_1, e_2, \cdots$ 独立同分布，其共同的分布为非正态的，且 $Ee_1=0, 0<Ee_1^2=\sigma^2<\infty$. 设当 n 充分大时 $S_n>0$. 则对 β_0 的 LSE $\hat{\beta}_n$ 而言，$S_n^{1/2}(\hat{\beta}_n-\beta_0)$ 渐近正态的充要条件为 $d_n\to 0$.

由于 LSE 有其简单的表达式：

$$S_n^{1/2}(\hat{\beta}_n-\beta_0)=S_n^{-1/2}\sum_{i=1}^n e_i x_i=\sum_{i=1}^n e_i x_{ni},$$

故上述命题的充分性容易直接由 Lindeberg 定理得出，或作为 M 估计的一般定理的一个推论. 为了证明必要性，我们需要下面的正态分解定理，它是由 P. Lévy 猜测并由 Cramér 证明了的(参看文献[45]中 p. 283～284).

引理 4.3　设 X_1 和 X_2 相互独立，且 X_1+X_2 服从正态分布，则 X_1 和 X_2 均服从正态分布.

我们需要指出，在此引理中，退化分布视为方差为 0 的正态分布.

下面证明定理 4.5 的必要性：假定 $S_n^{1/2}(\hat{\beta}_n-\beta_0)$ 渐近正态而 $d_n\equiv\max_{1\leqslant i\leqslant n}\|x_{ni}\|^2\not\to 0$，则存在 p 维单位向量 a，使 $\max_{1\leqslant i\leqslant n}|a'x_{ni}|\not\to 0$. 此时，$a'S_n^{1/2}(\hat{\beta}_n-\beta_0)=\sum_{i=1}^n e_i a'x_{ni}$ 是渐近正态的，但由 $\max_{1\leqslant i\leqslant n}|a'x_{ni}|\not\to 0$，$\sum_{i=1}^n e_i a'x_{ni}$ 的极限定律可以分解成两个成分之和，其中一个是非正态的，但由引理 4.3，这是不可能的. 这个矛盾就证明了 $d_n\to 0$.　▌

根据这个定理，我们有理由猜测，对一般的 M 估计的渐近正态性，在一定的附加条件下，条件 A4.5(即 $d_n\to 0$)也是必要的. 这一猜测还有待于证明.

下面我们用 $D(\Psi)$ 表示 Ψ 所有不连续点组成的集合. 一个重要的事实是：

命题 4.1　设 $G(u)=E\Psi(e_1+u)$ 在 $u=0$ 连续，则 $D(\Psi)$ 的 F 测度为 0.

证　当 $u>0$ 时，由 Ψ 的单调性，

$$\Psi(e_1+u)-\Psi(e_1-u)\geqslant 0.$$

故由 Fatou 引理，得

$$0=\lim_{u\to 0_+}(G(u)-G(-u))=\lim_{u\to 0_+}E[\Psi(e_1+u)-\Psi(e_1-u)]$$

$$\geqslant E\{\liminf_{u\to 0_+}[\Psi(e_1+u)-\Psi(e_1-u)]\}$$

$$=E[\Psi_+(e_1)-\Psi_-(e_1)].$$

由 $\Psi_+-\Psi_-\geqslant 0$，即得 $F\{D(\Psi)\}=0$. ▌

由此命题可知，若对 Ψ 的一种选取，A4.3 成立，则对此 Ψ 的选取，$F\{D(\Psi)\}=0$. 但对不同选取的 Ψ，所有的 $D(\Psi)$ 都是同一个集合，其 F 测度为 0. 对于不同选取的 Ψ，它们仅仅在 $D(\Psi)$ 上取不同的值，因而不会影响 $E\Psi(e_1+u)$、$E\Psi^2(e_1+u)$ 等函数的值.

最后我们指出，A4.4 中的条件

$$\lim_{u\to 0}E[\Psi(e_1+u)-\Psi(e_1)]^2=0$$

可用较弱的条件

$$\lim_{u\to 0}E\Psi^2(e_1+u)=E\Psi^2(e_1)$$

来代替. 确切地说，我们有以下命题：

命题 4.2　设 $F\{D(\Psi)\}=0$，且 $E\Psi^2(e_1)=\sigma^2<\infty$，则

$$\lim_{u\to 0}E[\Psi(e_1+u)-\Psi(e_1)]^2=0 \Leftrightarrow \lim_{u\to 0}E\Psi^2(e_1+u)=E\Psi^2(e_1).$$

证　只需证明"$\Leftarrow$"部分：由 $F\{D(\Psi)\}=0$ 知

$$\lim_{u\to 0}[\Psi(e_1+u)-\Psi(e_1)]^2=0,\quad \text{a.s.}.$$

因 $H(e_1,u)\equiv 2[\Psi^2(e_1+u)+\Psi^2(e_1)]-[\Psi(e_1+u)-\Psi(e_1)]^2\geqslant 0$，又

$$\lim_{u\to 0}H(e_1,u)=4\Psi^2(e_1),\quad \text{a.s.},$$

故由 Fatou 引理及 $\lim_{u\to 0}E\Psi^2(e_1+u)=E\Psi^2(e_1)$，有

$$4E\Psi^2(e_1)=E[\liminf_{u\to 0}H(e_1,u)]\leqslant \liminf_{u\to 0}EH(e_1,u)$$

$$=2\lim_{u\to 0}[E\Psi^2(e_1+u)+E\Psi^2(e_1)]$$

$$-\limsup_{u\to 0}E[\Psi(e_1+u)-\Psi(e_1)]^2$$

$$= 4E\Psi^2(e_1) - \limsup_{u\to 0} E[\Psi(e_1+u) - \Psi(e_1)]^2,$$

故

$$\limsup_{u\to 0} \quad E[\Psi(e_1+u) - \Psi(e_1)]^2 = 0. \qquad \blacksquare$$

§4.2 由估计方程定义的 M 估计的情形

在本节，我们只假定 $\hat{\beta}_n$ 满足(4.4)式，不必要求它满足估计方程(4.3)。为了建立 $\hat{\beta}_n$ 的渐近正态性，除了 A4.1 及 A4.3～A4.5 之外，我们引进下述假定以代替 A4.2：

A4.2* Ψ 为 R 上的非减函数.

我们有下面的定理：

定理 4.6 设在模型(4.1)之下，A4.1、A4.2*、A4.3～A4.5 均成立，$\hat{\beta}_n$ 满足(4.4)式. 则当 $n\to\infty$ 时，

$$S_n^{1/2}(\hat{\beta}_n - \beta_0) \xrightarrow{\mathscr{L}} N(0, \lambda^{-2}\sigma^2 I_p).$$

为了证明这个定理，我们可以直接利用引理 4.2 中的第二个结论. 但这个结论的证明用到了附录中所述的凸函数深入的分析性质，它们的证明虽然是现成的，但并不容易. 为此，我们宁愿用初等方法证明下面的引理：

引理 4.4 设 $\{a_{ni}, i=1,\cdots,n; n=1,2,\cdots\}$ 为一实数阵列，满足条件 $\sum_{i=1}^n a_{ni}^2 \leqslant 1 \quad (n=1,2,\cdots)$. 设 A4.1、A4.2*、A4.3～A4.5 成立，则对任意的常数 $C>0$，有

$$\sup_{|\gamma|\leqslant C} \left| \sum_{i=1}^n a_{ni}\{\Psi(e_i - x_{ni}'\gamma) - \Psi(e_i) + \lambda x_{ni}'\gamma\} \right| \xrightarrow{P} 0.$$

证 不失一般性，可设所有的 $a_{ni}\geqslant 0$. 记

$$\Psi_{ni}(\gamma) = \Psi(e_i - x_{ni}'\gamma) - \Psi(e_i) + \lambda x_{ni}'\gamma,$$

$$\Psi_{ni}^*(\gamma) = \Psi(e_i - x_{ni}'\gamma) - \Psi(e_i) - G(-x_{ni}'\gamma).$$

设 $\|\gamma_{i_1} - \gamma_{i_2}\| \leqslant \delta \quad (i=1,\cdots,n)$，则由 Schwarz 不等式，

$$\left|\sum_{i=1}^{n} a_{ni}\lambda x_{ni}'(\gamma_{i_1}-\gamma_{i_2})\right| \leqslant \lambda\sum_i a_{ni}\|x_{ni}\|\cdot\|\gamma_{i_1}-\gamma_{i_2}\|$$

$$\leqslant \lambda\delta\left(\sum_i a_{ni}^2\cdot\sum_i\|x_{ni}\|^2\right)^{1/2} \leqslant \lambda\delta\sqrt{p}. \qquad (4.19)$$

对于任给的 $\varepsilon>0$，我们可以把 $B=\{|\gamma|\leqslant C\}$ 分成若干个子集 B_1，$\cdots$，B_N，使每个 B_k 的直径 $<\varepsilon/(3\lambda\sqrt{p})$. 对每个 k，在 B_k 的闭包 $\overline{B}_k$ 中，存在 $\gamma_{i1}^{(k)}$ 和 $\gamma_{i2}^{(k)}$，它们可能与 n 有关，使得

$$-x_{ni}'\gamma_{i1}^{(k)}\leqslant -x_{ni}'\gamma\leqslant -x_{ni}'\gamma_{i2}^{(k)}，\text{对所有 } \gamma\in B_k.$$

由 Ψ 的单调性和(4.19)，我们有

$$\sup_{\gamma\in B_k}\left|\sum_{i=1}^{n} a_{ni}\Psi_{ni}(\gamma)\right| \leqslant \max_{j=1,2}\left|\sum_{i=1}^{n} a_{ni}\Psi_{ni}(\gamma_{ij}^{(k)})\right| + \varepsilon/3. \quad (4.20)$$

由 $d_n\to 0$，$\sup\limits_{1\leqslant i\leqslant n}|x_{ni}'\gamma_{ij}^{(k)}|\to 0\quad (j=1,2;k=1,\cdots,N)$. 由此，A4.3 及 $\sum\limits_{i=1}^{n}\|x_{ni}\|^2=p$，存在 ε_n，$0<\varepsilon_n\to 0$，使得当 n 充分大时，对 $j=1$，$2;k=1,\cdots,N$，有

$$\left|\sum_{i=1}^{n} a_{ni}(\Psi_{ni}(\gamma_{ij}^{(k)})-\Psi_{ni}^*(\gamma_{ij}^{(k)}))\right|$$

$$=\left|\sum_{i=1}^{n} a_{ni}(G(-x_{ni}'\gamma_{ij}^{(k)})+\lambda x_{ni}'\gamma_{ij}^{(k)})\right|$$

$$\leqslant \varepsilon_n\sum_{i=1}^{n} a_{ni}|x_{ni}'\gamma_{ij}^{(k)}| \leqslant \varepsilon_n\sum_i a_{ni}\|x_{ni}\|\cdot\|\gamma_{ij}^{(k)}\|$$

$$\leqslant c\sqrt{p}\,\varepsilon_n\left(\sum_i a_{ni}^2\sum_i\|x_{ni}\|^2\right)^{1/2} \leqslant cp\varepsilon_n. \quad (4.21)$$

故由(4.20)及(4.21)，当 n 充分大时，

$$P\left\{\sup_{|\gamma|\leqslant C}\left|\sum_{i=1}^{n} a_{ni}\Psi_{ni}(\gamma)\right|\geqslant\varepsilon\right\}$$

$$\leqslant P\left\{\sup_{1\leqslant k\leqslant N,j=1,2}\left|\sum_{i=1}^{n} a_{ni}\Psi_{ni}^*(\gamma_{ij}^{(k)})\right|\geqslant\varepsilon/2\right\}$$

$$\leqslant\sum_{k=1}^{N}\sum_{j=1}^{2}P\left\{\left|\sum_{i=1}^{n} a_{ni}\Psi_{ni}^*(\gamma_{ij}^{(k)})\right|\geqslant\varepsilon/2\right\}. \qquad (4.22)$$

另一方面，对所有的 j,k，由 A4.4 及 $\sup_{1\leqslant i\leqslant n}|x_{ni}'\gamma_{ij}^{(k)}|\to 0$，

$$\operatorname{Var}\left(\sum_{i=1}^{n}a_{ni}\Psi_{ni}^{*}(\gamma_{ij}^{(k)})\right)=\sum_{i=1}^{n}a_{ni}^{2}E[\Psi_{ni}^{*}(\gamma_{ij}^{(k)})]^{2}$$

$$\leqslant\sum_{i}a_{ni}^{2}E[\Psi(e_i-x_{ni}'\gamma_{ij}^{(k)})-\Psi(e_i)]^{2}\to 0,\text{当 } n\to\infty \text{ 时}.$$

由此式及(4.22)式即证明了引理.　　▌

现在我们来证明定理 4.6：

证　不失一般性，可设 $\beta_0=0$，令 $\beta(n)=S_n^{1/2}\hat{\beta}_n$，则由(4.4)，有

$$\left\|\sum_{i=1}^{n}\Psi(e_i-x_{ni}'\beta(n))x_{ni}\right\|=o_p(1).\tag{4.23}$$

往下证

$$|\beta(n)|=O_p(1).\tag{4.24}$$

我们指出，Rao 和 Zhao[56]在证明其定理 2.1 的过程中证得：若 A4.1、A4.2*、(2.1)～(2.3)成立，且 $E\Psi(e_1)=0$，则(4.23)式蕴涵了(4.24)式. 但在定理 4.6 的较强的条件下，我们可以给出一个简单得多的证明. 为此，设 U 是 R^p 中的单位球面，即

$$U=\{\beta\in R^p:\|\beta\|=1\}.$$

定义

$$D(\gamma,L)=\sum_{i=1}^{n}\Psi(e_i-Lx_{ni}'\gamma)x_{ni}'\gamma,\quad L\geqslant 0,\gamma\in R^p.$$

由 A 4.2*，对固定的 γ，$D(\gamma,L)$为 L 的非增函数.

任给 $\varepsilon>0$，取定

$$L_0\geqslant 8p^{1/2}\sigma/(\lambda\varepsilon^{1/2}).$$

由(4.23)式，存在 $n_1\geqslant n_0$，使

$$P\left\{\left|D\left(\frac{\beta(n)}{\|\beta(n)\|},\|\beta(n)\|\right)\right|\geqslant\frac{1}{4}L_0\lambda\right\}<\varepsilon/2,\text{当 } n\geqslant n_1.$$

由 $D(\gamma,\cdot)$的单调性，此式蕴涵了

$$P\{\|\beta(n)\|\geqslant L_0\}<P\left\{\sup_{\gamma\in U}D(\gamma,L_0)\geqslant-\frac{1}{4}L_0\lambda\right\}+\varepsilon/2.\tag{4.25}$$

由引理 4.4 及 $\sum_{i=1}^{n} x_{ni} x_{ni}' = I_p$，有

$$\sup_{\gamma \in U} \left| \sum_{i=1}^{n} [\Psi(e_i - L_0 x_{ni}' \gamma) - \Psi(e_i)] x_{ni}' \gamma + L_0 \lambda \right| \xrightarrow{P.} 0. \tag{4.26}$$

由 Schwarz 不等式，

$$\sup_{\gamma \in U} \left| \sum_{i=1}^{n} \Psi(e_i) x_{ni}' \gamma \right| \leqslant \left\| \sum_{i=1}^{n} \Psi(e_i) x_{ni} \right\|. \tag{4.27}$$

由(4.26)及(4.27)，存在 $n_2 \geqslant n_1$，当 $n \geqslant n_2$ 时，有

$$P\left\{ \sup_{\gamma \in U} D(\gamma, L_0) < -\frac{1}{2} L_0 \lambda + \left\| \sum_{i=1}^{n} x_{ni} \Psi(e_i) \right\| \right\} > 1 - \varepsilon/4. \tag{4.28}$$

又由 A4.4、$\sum_{i=1}^{n} \| x_{ni} \|^2 = p$ 及 L_0 的取法，有

$$P\left\{ \left\| \sum_{i=1}^{n} x_{ni} \Psi(e_i) \right\| \geqslant \frac{1}{4} L_0 \lambda \right\} \leqslant \left(\frac{4}{L_0 \lambda} \right)^2 E \left\| \sum_{i=1}^{n} x_{ni} \Psi(e_i) \right\|^2$$

$$= \left(\frac{4}{L_0 \lambda} \right)^2 \sum_{i=1}^{n} \sigma^2 \| x_{ni} \|^2 = 16 p \sigma^2 / (L_0 \lambda)^2 \leqslant \varepsilon/4. \tag{4.29}$$

由(4.28)及(4.29)，当 $n \geqslant n_2$ 时，

$$P\left\{ \sup_{\gamma \in U} D(\gamma, L_0) < -\frac{1}{4} L_0 \lambda \right\} > 1 - \varepsilon/2.$$

再由(4.25)式，当 $n \geqslant n_2$ 时，有

$$P\{ \| \hat{\beta}(n) \| \geqslant L_0 \} < \varepsilon/2 + \varepsilon/2 = \varepsilon. \tag{4.30}$$

至此(4.24)式得证. 由引理 4.4 及 $\sum_{i=1}^{n} x_{ni} x_{ni}' = I_p$，

$$I(\| \hat{\beta}(n) \| \leqslant L_0) \left| \sum_{i=1}^{n} \left[\Psi\left(e_i - x_{ni}' \hat{\beta}(n) \right) - \Psi(e_i) \right] x_{ni} + \lambda \hat{\beta}(n) \right| \xrightarrow{P.} 0.$$

此式与(4.30)及(4.23)式相结合，即得

$$\hat{\beta}(n) = \lambda^{-1} \sum_{i=1}^{n} x_{ni} \Psi(e_i) + o_p(1). \tag{4.31}$$

由 A4.1、A4.5 及 $E\Psi(e_1) = 0, 0 < E\Psi^2(e_1) = \sigma^2 < \infty$，Lindeberg 定

理适用于 $\sum_{i=1}^{n} x_{in}\Psi(e_i)$，即得定理 4.6. 证毕. ▌

以上的证明方法是综合了 Zhao 和 Chen[66] 以及 Rao 和 Zhao[56]两篇文献中所用的方法得到的.

§4.3 历史小记

关于回归系数 M 估计的渐近正态性的研究，较早的工作可以举 Relles（见文献[58]）、Huber（见文献[34]）和 Yohai（见文献[66]）的工作为代表. 他们分别研究了 p 固定和 p 随 n 趋于∞的情形. 在 Huber 的文章中，假定了以下三个条件：

1° 设计点列满足 Lindeberg 型的条件 A4.5；

2° ρ 是凸的，非单调的，且有充分高阶的有界导数，特别是 $\Psi(u)=\frac{\mathrm{d}}{\mathrm{d}u}\rho(u)$ 假定是连续有界的；

3° 误差序列 $\{e_i\}$ 是 iid. 序列，且 $E\Psi(e_1)=0$.

Huber 在 $d_n p^2\to 0$（这蕴涵 $p^3/n\to 0$）的条件下建立了回归系数 M 估计的渐近正态性.

1975 年，Bickel[19]研究了回归系数 β_0 的一步 M 估计的渐近正态性. 简言之，这种一步 M 估计是先从一个有理由认为较好的估计 β^* 出发，然后对估计方程(4.3)应用一步 Newton 方法以得出 β_0 的估计 β^{**}. 在某些条件下，β^* 与 β_0 的 M 估计 β_n 有相同的渐近正态性.

1979 年，Yohai 和 Maronna[67]系统研究了由估计方程(4.3)定义的回归系数 β_0 的 M 估计，分别在 p 固定和 p 随 n 趋于∞的情形，在一系列条件之下建立了估计的相合性和渐近正态性. 他们不需要 Ψ 是连续的，同时在 p 随 n 趋于∞的情形，也不需要 $d_n p^2\to 0$，只要求 $d_n p^{3/2}\to 0$（这蕴涵 $p^{3/2}/n\to 0$），这些条件都比 Huber 的有关条件要弱. 但他们的条件(A2)看来不大自然.

在 p 随 n 趋于∞的场合，关于由估计方程定义的回归系数的 M 估计的进一步研究，可以参看 Portnoy 的文献[51]和[52]以及

Welsh 的文献[64].

以下我们重点讨论 p 固定的场合.

在模型(4.14)中,设 $\hat{\beta}_R$ 是极值问题

$$\sum_{i=1}^{n} a_n(R_i)(Y_i - x_i'\beta) = \min \tag{4.32}$$

的解,此处 R_i 是 $\Delta i \equiv Y_i - x'_i\beta$ 在 $\Delta_1, \cdots, \Delta_n$ 中的秩,$a_n(\cdot)$是某个非减的得分函数,且 $\sum_{i=1}^{n} a_n(i) = 0$,则称 $\hat{\beta}_R$ 为模型(4.14)中回归系数的 R 估计(见文献[37]).但用这种方法不能估计模型(4.14)中的常数项 α_0. 1971 年,Jurečková[38]曾以另一种方式定义过 β_0 的 R 估计,并在较强的一组条件之下得到了 $\hat{\beta}_R$ 的渐近正态性. Jaeckel[37]在同样的条件下证明了这两种不同方式定义的 R 估计是渐近等价的.他还证明了(4.32)式左端是 β 的非负的(逐片线性的)凸函数.关于 M 估计和 R 估计之间的渐近关系,读者可参看文献[39]或文献[36].

1988 年,Heiler 和 Willers[30]利用凸函数的有关性质,建立了由(4.32)式定义的 β_0 的 R 估计 $\hat{\beta}_R$ 的渐近正态性.这种方法上的改进,不仅大大减弱了所加的条件,而且简化了证明.

1990 年,Chen、Bai、Zhao 和 Wu[4] 首次严格地建立了模型(4.1)之下 β_0 的最小一乘估计(LADE)和模型(4.14)中(α_0, β_0')的最小一乘估计的渐近正态性,即定理 4.3 和 4.4,所加条件也与这两个定理基本相同.同样的结果后来也为 Pollard[50]于 1991 年得到.

1990 年,Bai、Chen、Miao 和 Rao[11]建立了多元线性模型中最小距离估计的渐近性质.

1992 年,Bai、Rao 和 Wu[12]研究了下面很一般的多重线性回归模型:

$$Y_i = X_i'\beta_0 + \varepsilon_i \quad (i = 1, \cdots, n). \tag{4.33}$$

此处,Y_i 为第 i 次 p 维观测向量,X_i 为给定的 $m \times p$ 矩阵,β_0 为未知的 m 维回归系数向量,$\{\varepsilon_i\}$为一列 iid. p 维随机误差向量.他们

假定 ρ 是 p 个变量的给定的凸函数，并定义 β_0 的 M 估计 $\hat{\beta}_n$ 为使 $\sum_{i=1}^{n}\rho(Y_i-X_i'\beta)$ 取得最小值的任何 β 值. 他们利用凸函数的有关性质，在对 ρ、$\{X_i\}$ 和随机误差所加的尽可能弱的条件下，建立了 $\hat{\beta}_n$ 的渐近正态性，得出了有关线性假设的检验统计量的极限分布. 顺带着，他们也在关于 $\{X_i\}$ 的 Lindeberg 型的条件下建立了 $\hat{\beta}_n$ 的弱相合性(这个条件当然不是最弱的). 他们的渐近正态性的结果当 $p=1$ 时与定理 4.1 一致，并把文献[11]中研究的最小距离估计作为它的一个特例. 正如文献[12]中指出的，模型(4.33)比标准的多元线性回归模型

$$Y_i = B_0'x_i + \varepsilon_i \quad (i=1,\cdots,n) \tag{4.34}$$

更一般，此处 B_0 为 $m\times p$ 参数矩阵，$\{x_i\}$ 为 m 维向量的一个设计点列，Y_i、ε_i 与(4.33)中的一样.

1990 年，Koenker 和 Pontnoy[43]研究了重元回归的另一种意义下的 M 估计，它使得逐个坐标的残差的凸函数损失之和达到最小. 为求得渐近分布，他们也需要较强的条件.

关于由估计方程定义的 M 估计的渐近正态性的证明方法，是综合了 Zhao 和 Chen[68]及 Rao 和 Zhao[56]的两篇文章得到的. 后一篇文章证明了，若 A4.1、A4.2*、(2.1)～(2.3)成立，$E\Psi(e_1)=0$，且 $\hat{\beta}_n$ 满足(4.4)式，则 $\hat{\beta}_n$ 为模型(4.1)中 β_0 的一个弱相合估计. 这是与定理 2.1 平行的一个定理.

关于线性模型(4.1)中最小一乘估计的渐近正态性的研究，我们在这里要多费些笔墨来叙述. 证明 β_0 的最小一乘估计渐近正态性的首次努力是由 Bassett 和 Koenker 于 1978 年在文献[17]中作出的. 他们假定 $\{e_i\}$ 满足定理 4.3 中的条件及更强的条件，又假定极值问题(4.18)中的解唯一(这不是总能成立的，且难于验证)，以及

$$n^{-1}S_n \to Q,\ Q>0.$$

不幸的是，他们的论证包含了不易解决的严重错误(参看文献[4]).

Bloomfield 和 Steiger[20]及 Amemiya[7]也试图在他们各自的条件下建立 β_0 的最小一乘估计的渐近正态性，但正如文献[4]中指出的，他们的证明也存在着致命的错误.

当回归模型包含一常数项，即形如

$$Y_i = \alpha_0 + x'_i\beta_0 + e_i \quad (i = 1,\cdots,n)$$

(即(4.14)式)时，我们注意到 β_0 的 LADE $\hat{\beta}_n$ 是上文提到的由(4.32)式定义的 $\hat{\beta}_R$ 的一个特例(参看文献[20]中 p.62，引理 1).由此出发，可利用文献[38]、[37]及[30]中改进了的结果去证明 β_0 的 LADE $\hat{\beta}_n$ 的渐近正态性.但在文献[30]中，也假定了 F 有一个可微的密度 f 且具有有限的 Fisher 信息量，故 f 必须处处为正，甚至最简单的均匀分布也不满足这个条件.还有，正如前面指出的，它不能处理无常数项 α_0 的模型，而在带常数项的模型中，它也只能处理 β_0 的问题，而无法处理 α_0 的问题.

1987 年，Dupačová[28]证明了一个关于 β_0 的 LADE $\hat{\beta}_n$ 的渐近正态性的定理，其中假定$\{x_i\}$为随机观测值，她的结果用于参数无约束的情形时，大致与文献[20]中的相当，但包含了一个在数学上不理想的假定，即要求 $\|x_i\|$ 具有有限的三阶矩，且其方法不能处理 x_i 非随机这个更重要的情况.

附录　凸函数的有关性质

为便于读者参考，我们在这个附录中不加证明地罗列出关于凸分析的若干知识，在第 4、5 章中要用到.对这些性质的证明以及关于凸分析的比较系统的论述，读者可参看 Rockafellar 的专著[59]及 Heiler 和 Willers 的论文[30].

首先从介绍凸集的定义开始.称 $A\subset R^p$ 为一个凸集，若对任何u、$v\in A$及任何 $\lambda\in(0,1)$，总有

$$\lambda u + (1-\lambda)v \in A.$$

即一个凸集若包含某两点 u 和 v，则必包含连接 u 和 v 的整个线

段. 由此易见，若 A 为凸集，$u_i\in A$，$\lambda_i\geqslant 0$ $(i=1,\cdots,n)$ 且 $\sum_{i=1}^{n}\lambda_i=1$，则

$$\sum_{i=1}^{n}\lambda_i u_i\in A,$$

其中 $n\geqslant 2$ 为任意整数. 称 $\sum_{i=1}^{n}\lambda_i u_i$ 为 $u_1,\cdots,u_n$ 的一个凸组合.

设 D 为 R^p 中的一个非空凸子集，称函数 $\rho: D\to R^1$ 为 D 上定义的一个凸函数，若对任何 u、$v\in D$ 及任何 $\lambda\in(0,1)$，总有

$$\rho(\lambda u+(1-\lambda)v)\leqslant\lambda\rho(u)+(1-\lambda)\rho(v).$$

称 ρ 为严格凸的，若对上述的任何 u、v 及 λ，总有严格的不等式成立：

$$\rho(\lambda u+(1-\lambda)v)<\lambda\rho(u)+(1-\lambda)\rho(v).$$

由此易见，若 ρ 为 D 上的凸函数，$u_i\in D$，$\lambda_i\geqslant 0$ $(i=1,\cdots,n)$，且 $\sum_{i=1}^{n}\lambda_i=1$，则

$$\rho\left(\sum_{i=1}^{n}\lambda_i u_i\right)\leqslant\sum_{i=1}^{n}\lambda_i\rho(u_i).$$

这就是熟知的 Jensen 不等式.

以下我们总假定 D 为 R^p 中的非空开凸集，且定义于 D 上的凸函数 ρ 总是取有限值. 在下面，$\langle\cdot,\cdot\rangle$ 总表示 R^p 中通常的内积，$\|\cdot\|$ 表示欧氏模. 我们将不加证明地引述凸函数的一些重要性质如下：

1) 开凸集 D 上的凸函数 ρ 必为 D 上的连续函数. 且对 D 的任一紧子集 S，ρ 在 S 上满足 Lipschitz 条件，即存在常数 $K>0$，使得对任何 u、$v\in S$，有

$$|\rho(u)-\rho(v)|\leqslant K\|u-v\|.$$

2) 设 $\{\rho_i, i\in I\}$ 为在开凸集 D 上定义的一族凸函数，I 为任意的指标集. 假定

a) 存在 D 的一稠密子集 D'，使 $\sup\{\rho_i(u), i\in I\}$ 对每个 $u\in D'$ 为有限的；

b) 至少存在一个 $u\in D$，使 $\inf\{\rho_i(u), i\in I\}$ 是有限的.

则对 D 的每个紧子集 S，$\{\rho_i, i\in I\}$在 S 上满足等度 Lipschitz 条件，即存在正的常数 K，使对任意的 u、$v\in S$，任意的 $i\in I$，有

$$|\rho_i(u)-\rho_i(v)|\leqslant K\|u-v\|.$$

3) 设 D 为一个开凸集，$\rho_1,\rho_2,\cdots$为 D 上的一列凸函数. 假定存在 D 的一个稠密子集 D'，使得对每个 $u\in D'$，$\rho_1(u),\rho_2(u),\cdots$都有有限的极限，则对每个 $u\in D$，极限

$$\rho(u)=\lim_{n\to\infty}\rho_n(u)$$

存在有限，且函数 $\rho(\cdot)$在 D 上是凸的. 进一步，上式中的收敛在 D 的每个紧子集 S 上都是一致的.

设 f 为 R^p 上的实值函数，$u\in R^p$，v 为一个给定的 p 维向量. 若极限

$$f'(u;v)=\lim_{\lambda\downarrow 0}\frac{f(u+\lambda v)-f(u)}{\lambda}$$

存在(可以为$+\infty$或$-\infty$)，则称此极限为 f 在 u 关于向量 v 的*单边方向导数*. 注意到

$$-f'(u;-v)=\lim_{\lambda\uparrow 0}\frac{f(u+\lambda v)-f(u)}{\lambda},$$

故单边方向导数 $f'(u;v)$是双边的，当且仅当 $f'(u;-v)$存在，且

$$f'(u;-v)=-f'(u;v).$$

称 f 在 u *是可微的*，若存在一个 p 维向量 u^*，使

$$\lim_{z\to u}\{f(z)-f(u)-\langle u^*,z-u\rangle\}/\|z-u\|=0.$$

这样的 u^* 如果存在，则称它为 f 在 u 的*梯度*(gradient)，记为 $\nabla f(u)$. 显然，若 f 在 u 可微，则 f 在 u 的所有方向导数都是有限且双边的，且

$$f'(u;v)=\langle\nabla f(u),v\rangle,\text{对任意的 } v\in R^p.$$

设 ρ 为在凸集 D 上定义的一个凸函数，向量 u^* 称为 ρ 在点 u 处的一个*次梯度*(subgradient). 如果

$$\rho(z)\geqslant\rho(u)+\langle u^*,z-u\rangle,\text{对任意的 } z\in R^p.$$

此时由方程

$$f(z)=\rho(u)+\langle u^*,z-u\rangle\quad(z\in R^p)$$

定义的 R^{p+1} 中的超平面 $\{(z,f(z)),z\in R^p\}$ 称为凸函数 ρ 在 u 的一个支撑平面，它通过点 $(u,\rho(u))$ 且位于 ρ 的图象的下方.

ρ 在 u 的所有次梯度所成的集合称为 ρ 在 u 的次微分(subdifferential)，记为 $\partial\rho(u)$. 多值映照 $\partial\rho:u\mapsto\partial\rho(u)$ 称为函数 ρ 的次微分. 很明显，$\partial\rho(u)$ 是一个闭凸集. 一般地说，$\partial\rho(u)$ 可能是空集，也可能由唯一的一个向量组成. 如果 $\partial\rho(u)$ 是非空的，则称 ρ 在点 u 处是次可微的.

4) 设 ρ 为开凸集 D 上的凸函数，$u\in D$. 对每个 v，在定义 $\rho'(u;v)$ 时的差商 $[\rho(u+\lambda v)-\rho(u)]/\lambda$ 是 $\lambda>0$ 的非减函数，因而 $\rho'(u;v)$ 总存在：

$$\rho'(u;v)=\inf_{\lambda>0}\frac{\rho(u+\lambda v)-\rho(u)}{\lambda},$$

且 $$-\rho'(u;-v)\leqslant\rho'(u;v),\text{对任意的 } v\in R^p.$$

对每个 $u\in D$，$\partial\rho(u)$ 为一非空有界(闭凸)集，且 $u^*\in\partial\rho(u)$ 当且仅当 $\rho'(u;v)\geqslant\langle u^*,v\rangle$，对任意的 $v\in R^p$.

5) 设 $\rho_1,\rho_2,\cdots$ 为开凸集 D 上的一列凸函数，且在 D 上点点收敛到 D 上的一个凸函数 ρ. 设 $u\in D$，$u_1,u_2,\cdots$ 为 D 中的一个点列，收敛到 u. 则对任给的 $\varepsilon>0$，存在 n_0，当 $n\geqslant n_0$ 时，

$$\partial\rho_i(u_i)\subset\partial\rho(u)+\varepsilon B,$$

其中 B 为 R^p 中的 Euclidean 单位球，$\varepsilon B=\{\varepsilon u,u\in B\}$.

特别是，对任何 $u\in D$ 及任给的 $\varepsilon>0$，存在 $\delta>0$，使

$$\partial\rho(z)\subset\partial\rho(u)+\varepsilon B,\text{对任意的 } z\in u+\delta B.$$

6) 凸函数 ρ 的次微分 $\partial\rho$ 具有以下的循环单调性(cyclical monotonity)，即对任给的 n，任意的 (u_i,u_i^*)，其中 $u_i^*\in\partial\rho(u_i)$ $(i=1,\cdots,n)$，有

$$\langle u_1-u_0,u_0^*\rangle+\langle u_2-u_1,u_1^*\rangle+\cdots+\langle u_n-u_{n-1},u_{n-1}^*\rangle+\langle u_0-u_n,u_n^*\rangle\leqslant 0$$

这是因为，由次梯度的定义，

$$\text{上式左方}\leqslant[\rho(u_1)-\rho(u_0)]+[\rho(u_2)-\rho(u_1)]+\cdots+[\rho(u_n)-\rho(u_{n-1})]+[\rho(u_0)-\rho(u_n)]=0.$$

7）设 ρ 为开凸集 D 上的凸函数.若 ρ 在 $u\in D$ 是可微的，则 $\nabla\rho(u)$ 是 ρ 在 u 点唯一的次梯度，故

$\rho(z)\geqslant\rho(u)+\langle\nabla\rho(u),z-u\rangle$，对任意的 $z\in D$.

反之，若 ρ 在 $u\in D$ 有唯一的次梯度，则 ρ 在 u 点是可微的.

8）设 ρ 是在开凸集 D 上定义的凸函数，D_1 是函数 ρ 可微的点所成的集合.则 D_1 在 D 中稠密，且差集 $D\backslash D_1$ 是一个 Lebesgue 零测集.进一步，梯度映照 $\nabla\rho:u\mapsto\nabla\rho(u)$ 限于 D_1 上是连续的.

特别是，若 ρ 在 D 上可微（即在 D 上点点可微），则 ρ 必为连续可微的.

可以证明，D_1 是一个 G_δ 型的集合，即 D_1 是至多可列个开集的交集.

9）设 ρ 为开凸集 D 上可微凸函数，$\rho_1,\rho_2,\cdots$ 为 D 上一列凸函数，$\lim\limits_{n\to\infty}\rho_n(u)=\rho(u)$ 对任意的 $u\in D$.则对任意的 $\Psi_n(u)\in\partial\rho_n(u)$，有

$$\lim_{n\to\infty}\Psi_n(u)=\nabla\rho(u),\text{对任意的 } u\in D,$$

而且在 D 的任一紧子集 S 上，上述收敛是一致的.

10）设 $\rho,\rho_1,\rho_2,\cdots$ 都是开集 D 上的凸函数，ρ 在 D 上可微，D' 为 D 的一个稠密子集，$\Psi_n(u)\in\partial\rho_n(u)$，且

$$\lim_{n\to\infty}\Psi_n(u)=\nabla\rho(u),\text{对任意的 } u\in D',$$

$$\lim_{n\to\infty}\rho_n(u_0)=\rho(u_0),\text{对至少一个 } u_0\in D',$$

则

$$\lim_{n\to\infty}\rho_n(u)=\rho(u),\text{对任意的 } u\in D,$$

且上述收敛在 D 的任一紧子集 S 上是一致的.

对于一元凸函数，有

11）设 ρ 为开区间 (a,b) 上的凸函数，Ψ_-,Ψ_+ 分别为 ρ 的左、右导函数，则 Ψ_-,Ψ_+ 在 (a,b) 上都是非减的，有限的，且

$\Psi_-(z_1)\leqslant\Psi_-(u)\leqslant\Psi_+(u)\leqslant\Psi_-(z_2)$，当 $a<z_1<u<z_2<b$.

进一步，对每个 $u\in(a,b)$，

$$\lim_{z\downarrow u}\Psi_+(z)=\Psi_+(u),\ \lim_{z\uparrow u}\Psi_-(z)=\Psi_-(u),$$

$$\lim_{z \downarrow u} \Psi_-(z) = \Psi_+(u), \quad \lim_{z \uparrow u} \Psi_+(z) = \Psi_-(u).$$

若记

$$D_1 = \{u \in (a,b), \rho \text{ 在 } u \text{ 可微}\},$$
$$D_+ = \{u \in (a,b), \Psi_+ \text{ 在 } u \text{ 连续}\},$$
$$D_- = \{u \in (a,b), \Psi_- \text{ 在 } u \text{ 连续}\}.$$

则 $D_1 = D_+ = D_-$，且$(a,b) \backslash D_1$至多为一可列集.

进一步，设 Ψ 为(a,b)上的任一实函数，满足

$$\Psi_-(u) \leqslant \Psi(u) \leqslant \Psi_+(u), \text{对任意的 } u \in (a,b),$$

则对任给的 $u_0 \in (a,b)$，有

$$\rho(u) = \rho(u_0) + \int_{u_0}^{u} \Psi(v) \mathrm{d}v.$$

在结束本附录之前，我们指出，上面的性质9)源于文献[59]中的定理25.7. Rockafellar 要求所有的 ρ_n 在 D 上都是可微的. 但正如 Heiler 和 Willers 在文献[30]中所指出的，在 Rockafellar 的证明中，并没有用到 ρ_n 在 D 上的可微性，只用到了极限函数 ρ 在 D 上的连续可微性质，因此性质9)可以在上面所说的较弱的条件下成立. 上面这些性质的证明，除性质10)之外，可参看文献[59]中的第10节及第23～25节；性质10)的证明可参看文献[30].

第 5 章

M 检验统计量的渐近理论

考虑通常的线性回归模型

$$Y_i = x_i'\beta + e_i \quad (i = 1,\cdots,n), \tag{5.1}$$

此处 Y_i 为第 i 次观测值，$\{x_i\}$为已知的 p 维试验点列，β 为未知的 p 维回归参数向量，$\{e_i\}$为试验误差序列. 在本章中我们假定 §4.1 中的条件 A4.1～A4.5 成立.

在模型(5.1)之下，我们常常要考虑以下的假设检验问题：

$$H_0: H'(\beta - b) = 0 \leftrightarrow H_1: H'(\beta - b) \neq 0, \tag{5.2}$$

此处 H 为已知的秩为 q 的 $p\times q$ 矩阵，b 为已知的 p 维向量($0<q<p$)，令

$$\inf_{H'(\beta-b)=0}\sum_{i=1}^{n}\rho(Y_i - x_i'\beta) \equiv \sum_{i=1}^{n}\rho(Y_i - x_i'\tilde{\beta}_n) \equiv \tilde{M}_n, \tag{5.3}$$

$$\inf_{\beta\in R^p}\sum_{i=1}^{n}\rho(Y_i - x_i'\beta) \equiv \sum_{i=1}^{n}\rho(Y_i - x_i'\hat{\beta}_n) \equiv \hat{M}_n. \tag{5.4}$$

为了检验线性假设(5.2)，我们可以使用 $\tilde{M}_n-\hat{M}_n$ 作为衡量数据与原假设偏离程度的一个准则. 我们不妨将此准则称为 M 准则. 如果能够求出它在 H_0 之下的极限分布，并找出多余参数的相合估计，即可构造出需要的检验统计量和有关的检验.

为了检验假设(5.2)，我们也可以使用其他的检验准则. 比较常用的有 Wald 检验准则 W_n，Rao 的计分型准则(Score type criterion，参看文献[53])R_n 以及它的一种变化形式 R_n^*(参看文献[61])：

$$W_n = (\hat{\beta}_n - b)'H(H'S_n^{-1}H)^{-1}H'(\hat{\beta}_n - b)$$
$$= \hat{\beta}(n)'H_nH_n'\hat{\beta}(n), \tag{5.5}$$
$$R_n = \Psi_n(\tilde{\beta}_n)'\Psi_n(\tilde{\beta}_n), \tag{5.6}$$
$$R_n^* = \Psi_n(\tilde{\beta}_n)'H_nH'_n\Psi_n(\tilde{\beta}_n), \tag{5.7}$$

以上诸式中

$$\hat{\beta}(n) = S_n^{1/2}(\hat{\beta}_n - b), \tag{5.8}$$
$$H_n = S_n^{-1/2}H(H'S_n^{-1}H)^{-1/2}: p \times q, \tag{5.9}$$
$$\Psi_n(\tilde{\beta}_n) = S_n^{-1/2}\sum_{i=1}^{n}x_i\Psi(Y_i - x_i'\tilde{\beta}_n), \tag{5.10}$$

其中 H_n 满足

$$H_n'H_n = I_q. \tag{5.11}$$

在下面 § 5.1 中我们将求出上面这些检验准则在一列局部对立假设

$$H_{2,n}: H'(\beta - b) = H'\omega_n \tag{5.12}$$

之下的极限分布，其中 p 维向量 ω_n 满足

$$\| S_n^{1/2}\omega_n \| = O(1)，当\ n \to \infty. \tag{5.13}$$

特别是，当 $\omega_n = 0$ 时即得到它们在原假设 H_0 之下的极限分布. 在 § 5.1 中我们还将构造出多余参数 λ 及 σ^2 的相合估计，从而得出有关的检验统计量及其极限分布.

在 § 5.2 中我们将考虑通常的多重线性模型

$$Y_i = B'x_i + \varepsilon_i \quad (i = 1, \cdots, n), \tag{5.14}$$

其中 Y_i 为 p 维观测值，ε_i 为 p 维随机误差向量，$\{x_i\}$ 为已知的 m 维设计点列，B 为未知的 $m \times p$ 回归参数矩阵. 在该节中我们将研究在模型(5.14)之下有关的线性假设的各种检验准则的极限分布，并研究多余参数的估计问题. 在 § 5.3 中我们将简略地介绍一些有关的文献.

§ 5.1　简单模型中的渐近理论

在线性回归模型(5.1)之下，为了检验线性假设(5.2)，我们提

出了若干检验准则.本节将在A4.1～A4.5的条件之下，求出它们在局部对立假设(5.12)之下的极限分布，并构造出多余参数 λ 和 σ^2 的相合估计 $\hat{\lambda}_n$ 和 $\hat{\sigma}_n^2$.

首先，我们将模型(5.1)写成规范形式.为此，令 $S_n = \sum_{i=1}^n x_i x_i'$.由A4.5及 $n \geqslant n_0$ 的约定，S_n 为 p 阶正定阵.令

$$x_{ni} = S_n^{-1/2} x_i,\quad \beta(n) = S_n^{1/2}(\beta - b),$$
$$Y_{ni} = Y_i - x_i' b. \tag{5.15}$$

则模型(5.1)可以写成以下规范形式：

$$Y_{ni} = x_{ni}'\beta(n) + e_i \quad (i = 1,\cdots,n), \tag{5.16}$$

其中

$$\sum_{i=1}^n x_{ni} x_{ni}' = I_p, \tag{5.17}$$

$$d_n \equiv \max_{1 \leqslant i \leqslant n} \| x_{ni} \|^2 \to 0, \text{当 } n \to \infty. \tag{5.18}$$

我们有以下的定理：

定理5.1　设在模型(5.1)之下，A4.1～A4.5成立，且真参数 β 满足(5.12)及(5.13).则当 $n \to \infty$ 时，

$$\begin{aligned} 2\lambda\sigma^{-2}(\tilde{M}_n - \hat{M}_n) &= \lambda^2\sigma^{-2}W_n + o_p(1) \\ &= \sigma^{-2}R_n + o_p(1) = \sigma^{-2}R_n^* + o_p(1) \\ &= \left\| \sigma^{-1}\sum_{i=1}^n H_n' x_{ni} \Psi(e_i) + \sigma^{-1}\lambda\delta(n) \right\|^2 + o_p(1), \end{aligned}$$

$$\tag{5.19}$$

其中 H_n、x_{ni} 分别参看(5.9)和(5.15)，而

$$\delta(n) = H_n' S_n^{1/2} \omega_n : q \times 1. \tag{5.20}$$

在定理5.1的条件下，由Lindeberg定理，有

$$\sigma^{-1}\sum_{i=1}^n H_n' x_{ni} \Psi(e_i) \xrightarrow{\mathscr{L}} N(0, I_q), \tag{5.21}$$

因而当参数与原假设 H_0 偏离不大时(在(5.12)和(5.13)式的意义之下)，$2\lambda\sigma^{-2}(\tilde{M}_n - \hat{M}_n)$、$\lambda^2\sigma^{-2}W_n$、$\sigma^{-2}R_n$ 及 $\sigma^{-2}R_n^*$ 都渐近等价于自由度为 q，非中心参数为 ν_n 的非中心 χ^2 分布 χ^2_{q,ν_n}，其中

$$\nu_n = \lambda^2\sigma^{-2}\delta'(n)\delta(n) = \lambda^2\sigma^{-2}\omega_n' S_n^{1/2} H_n H_n' S_n^{1/2}\omega_n$$
$$= \lambda^2\sigma^{-2}\omega_n' H(H'S_n^{-1}H)^{-1}H'\omega_n. \quad (5.22)$$

特别是，若当 $n\to\infty$ 时，$\| S_n^{1/2}\omega_n \| \to 0$，则

$$2\lambda\sigma^{-2}(\widetilde{M}_n - \hat{M}_n) = \lambda^2\sigma^{-2}W_n + o_p(1) = \sigma^{-2}R_n + o_p(1)$$
$$= \sigma^{-2}R_n^* + o_p(1) \xrightarrow{\mathscr{L}} \chi_q^2, \quad (5.23)$$

即极限分布为自由度为 q 的中心 χ^2 分布.

为了估计 σ^2，我们建议使用估计量

$$\hat{\sigma}_n^2 = n^{-1}\sum_{i=1}^n \Psi^2(Y_i - x_i'\hat{\beta}_n) = n^{-1}\sum_{i=1}^n \Psi^2(Y_{ni} - x_{ni}'\hat{\beta}(n)). \quad (5.24)$$

如果$\frac{\mathrm{d}}{\mathrm{d}u}E\Psi(e_1+u)\Big|_{u=0} = E\frac{\mathrm{d}}{\mathrm{d}u}\Psi(e_1+u)\Big|_{u=0}$ 成立，也就是说，若 $\lambda = E\frac{\mathrm{d}\Psi}{\mathrm{d}u}(e_1)$，我们也可用

$$n^{-1}\sum_{i=1}^n \frac{\mathrm{d}\Psi}{\mathrm{d}u}(Y_{ni} - x_{ni}'\hat{\beta}(n))$$

作为 λ 的一个估计，然后再去建立它的相合性. 但是，即使在 $\rho(u) = |u|$ 这种最简单的情形，上述条件也不成立. 因此，我们必须寻求 λ 的更一般的估计. 为此，可取 $h=h_n>0$，使

$$h_n/d_n^{1/3} \to \infty,\ h_n \to 0,\ 且\lim_{n\to\infty}\inf nh_n^2 > 0, \quad (5.25)$$

然后我们可用

$$\hat{\lambda}_n = (2nh)^{-1}\sum_{i=1}^n \{\Psi(Y_i - x_i'\hat{\beta}_n + h)$$
$$- \Psi(Y_i - x_i'\hat{\beta}_n - h)\} \quad (5.26)$$

作为 λ 的一个估计. 我们有下述定理：

定理 5.2　设在模型(5.1)之下，A4.1～A4.5 及(5.25)均满足. 则

$$\lim_{n\to\infty}\hat{\sigma}_n^2 = \sigma^2,\ \text{in pr.}, \quad (5.27)$$
$$\lim_{n\to\infty}\hat{\lambda}_n = \lambda,\ \text{in pr.} \quad (5.28)$$

因此，当我们用 $\hat{\lambda}_n$ 代替 λ，用 $\hat{\sigma}_n^2$ 代替 σ^2 时，即可得到有关的检

验统计量 $2\hat{\lambda}_n\hat{\sigma}_n^{-2}(\widetilde{M}_n-\hat{M}_n)$、$\hat{\lambda}_n^2\hat{\sigma}_n^{-2}W_n$、$\hat{\sigma}_n^{-2}R_n$ 及 $\hat{\sigma}_n^{-2}R_n^*$，它们也有类似于(5.19)的渐近等价表达式. 当 $\| S_n^{1/2}\omega_n \| \to 0$ 时，它们也有类似于(5.23)式的自由度为 q 的中心 χ^2 极限分布.

为了证明定理 5.1，设 K 是一个秩为 $p-q$ 的 $p\times(p-q)$ 矩阵，使得

$$H'K = 0. \tag{5.29}$$

不失一般性，可设 $K'\omega_n=0$. 则 H_0 和 $H_{2,n}$ 可重写为：

$$H_0:\beta - b = K\gamma, \text{对某个 } \gamma \in R^{p-q},$$

$$H_{2,n}:\beta - b = K\gamma + \omega_n, \text{对某个 } \gamma \in R^{p-q}.$$

我们已在(5.9)中定义 $H_n=S_n^{-1/2}H(H'S_n^{-1}H)^{-1/2}$，今定义 $K_n=S_n^{1/2}K(K'S_nK)^{-1/2}$. 则

$$H_n'H_n = I_q,\ K_n'K_n = I_{p-q},\ H_n'K_n = 0. \tag{5.30}$$

令 $\gamma_0(n)=(K'S_nK)^{1/2}\gamma$，则当 H_0 成立时，模型(5.16)可写为

$$Y_{ni} = x_{ni}'K_n\gamma_0(n) + e_i \quad (i = 1,\cdots,n). \tag{5.31}$$

令 $\delta(n)=H_n'S_n^{1/2}\omega_n$（见(5.20)式），且令

$$\gamma(n) = (K'S_nK)^{1/2}\gamma + K_n'S_n^{1/2}\omega_n. \tag{5.32}$$

由(5.30)，$K_nK_n'+H_nH_n'=I_p$. 由此式及(5.29)式，在 $H_{2,n}$ 之下，

$$\beta(n) = K_n\gamma(n) + H_n\delta(n). \tag{5.33}$$

在(5.8)中我们已经定义 $\hat{\beta}(n)=S_n^{1/2}(\hat{\beta}_n-b)$，易知它即为模型(5.16)之下 $\beta(n)$ 的 M 估计. 类似地，设 $\hat{\gamma}(n)$ 为模型(5.31)之下 $\gamma_0(n)$ 的 M 估计，即 $\hat{\gamma}(n)$ 满足条件

$$\sum_{i=1}^n\rho(Y_{ni} - x_{ni}'K_n\hat{\gamma}(n)) = \inf_{\zeta\in R^{p-q}}\sum_{i=1}^n\rho(Y_{ni} - x_{ni}'K_n\zeta) = \widetilde{M}_n. \tag{5.34}$$

由此容易得出

$$\widetilde{\beta}_n = b + S_n^{-1/2}K_n\hat{\gamma}(n). \tag{5.35}$$

由引理 4.2 容易得到：

引理 5.1 设在模型(5.1)之下，A4.1～A4.5、(5.12)式及(5.13)式均满足，则对任何常数 $C>0$，及由(5.33)式定义的 $\beta(n)$ 和(5.32)式定义的 $\gamma(n)$，

$$\sup_{|\xi-\beta(n)|\leqslant C}\left|\sum_{i=1}^{n}\{\rho(Y_{ni}-x_{ni}'\xi)-\rho(e_i)\}+\sum_{i=1}^{n}\Psi(e_i)x_{ni}'(\xi-\beta(n))\right.$$
$$\left.-\frac{\lambda}{2}\|\xi-\beta(n)\|^2\right|\xrightarrow{P.}0,$$

$$\sup_{|\zeta-\gamma(n)|\leqslant C}\left|\sum_{i=1}^{n}[\rho(Y_{ni}-x_{ni}'K_n\zeta)-\rho(e_i)]\right.$$
$$+\sum_{i=1}^{n}\Psi(e_i)x_{ni}'(K_n\zeta-\beta(n))$$
$$\left.-\frac{\lambda}{2}(\|\zeta-\gamma(n)\|^2+\|\delta(n)\|^2)\right|\xrightarrow{P.}0,$$

且

$$\sup_{|\xi-\beta(n)|\leqslant C}\left|\sum_{i=1}^{n}[\Psi(Y_{ni}-x_{ni}'\xi)-\Psi(e_i)]x_{ni}+\lambda(\xi-\beta(n))\right|\xrightarrow{P.}0,$$

$$\sup_{|\zeta-\gamma(n)|\leqslant C}\left|\sum_{i=1}^{n}[\Psi(Y_{ni}-x_{ni}'K_n\zeta)-\Psi(e_i)]x_{ni}+\lambda(K_n\zeta-\beta(n))\right|\xrightarrow{P.}0.$$

证　以第二式的证明为例，记 $\zeta-\gamma(n)=\eta$，它等价于

$$\sup_{|\eta|\leqslant C}\left|\sum_{i=1}^{n}\{\rho(e_i-x_{ni}'[K_n\eta-H_n\delta(n)])-\rho(e_i)\}\right.$$
$$+\sum_{i=1}^{n}\Psi(e_i)x_{ni}'(K_n\eta-H_n\delta(n))$$
$$\left.-\frac{\lambda}{2}(\|\eta\|^2+\|\delta(n)\|^2)\right|\xrightarrow{P.}0.$$

由(5.13)和(5.20)，当 $|\eta|\leqslant C$ 时，

$$|K_n\eta-H_n\delta(n)|\leqslant C_1\quad(C_1\text{ 为一常数}).$$

由引理 4.2，注意到(5.30)，即得所要的结果.　▍

引理 5.2　设在线性模型(5.1)之下，A4.1～A4.5、(5.12)式及(5.13)式均满足. 则当 $n\to\infty$ 时，

$$\hat{\beta}(n)-\beta(n)=\lambda^{-1}\sum_{i=1}^{n}\Psi(e_i)x_{ni}+o_p(1),$$

$$\hat{\gamma}(n)-\gamma(n)=\lambda^{-1}\sum_{i=1}^{n}K_n'x_{ni}\Psi(e_i)+o_p(1),$$

以上式中 $\beta(n)=K_n\gamma(n)+H_n\delta(n)$.

证 以第二式的证明为例，令

$$\bar{\gamma}(n) - \gamma(n) = \lambda^{-1}\sum_{i=1}^{n} K_n' x_{ni}\Psi(e_i), \tag{5.36}$$

由 $d_n = \max_{1\leqslant i\leqslant n} \| x_{ni} \|^2 \to 0$ 及有关的条件，知 Lindeberg 条件成立，故上式有渐近的多维正态分布，故

$$|\bar{\gamma}(n) - \gamma(n)| = O_p(1), \text{当 } n \to \infty. \tag{5.37}$$

由引理 5.1，对任何 $c>0$ 及 $\delta>0$，

$$\sup_{\|\zeta-\bar{\gamma}(n)\|=\delta} I(\|\bar{\gamma}(n) - \gamma(n)\| \leqslant C)\left|\sum_{i=1}^{n}\rho(Y_{ni} - x_{ni}'K_n\zeta) - \sum_{i=1}^{n}\rho(Y_{ni} - x_{ni}'K_n\bar{\gamma}(n)) - \frac{\lambda}{2}\|\zeta - \bar{\gamma}(n)\|^2\right| \xrightarrow{P.} \tag{5.38}$$

故由与定理 4.1 中类似的论证，有

$$P\left\{\inf_{\|\zeta-\bar{\gamma}(n)\|=\delta}\sum_{i=1}^{n}\rho(Y_{ni} - x_{ni}'K_n\zeta) \geqslant \sum_{i=1}^{n}\rho(Y_{ni} - x_{ni}'K_n\bar{\gamma}(n)) + \frac{\lambda}{4}\delta^2\right\} \to 1. \tag{5.39}$$

由此及 ρ 的凸性，可得

$$P\{\|\hat{\gamma}(n) - \bar{\gamma}(n)\| \geqslant \delta\} \to 0, \text{当 } n \to \infty.$$

由 δ 的任意性，

$$\hat{\gamma}(n) - \bar{\gamma}(n) \xrightarrow{P.} 0. \tag{5.40}$$

引理 5.2 即由(5.36)式及(5.40)式得出. ▌

现在我们来证明定理 5.1：

证 由引理 5.2，

$$\|\hat{\beta}(n) - \beta(n)\| = O_p(1). \tag{5.41}$$

由引理 5.1 及(5.41)式，

$$\sum_{i=1}^{n}[\rho(Y_{ni} - x_{ni}'\hat{\beta}(n)) - \rho(e_i)] + \sum_{i=1}^{n}\Psi(e_i)x_{ni}(\hat{\beta}(n) - \beta(n)) - \frac{\lambda}{2}\|\hat{\beta}(n) - \beta(n)\|^2 \xrightarrow{P.} 0.$$

再由引理 5.2，

$$\sum_{i=1}^{n}[\rho(Y_{ni}-x_{ni}'\hat{\beta}(n))-\rho(e_i)]+\frac{1}{2\lambda}\left\|\sum_{i=1}^{n}\Psi(e_i)x_{ni}\right\|^2\xrightarrow{P.}0. \tag{5.42}$$

同样可得

$$\sum_{i=1}^{n}[\rho(Y_{ni}-x_{ni}'K_n\hat{\gamma}(n))-\rho(e_i)]+\frac{1}{2\lambda}\left\|\sum_{i=1}^{n}K_n'x_{ni}\Psi(e_i)\right\|^2 - \sum_{i=1}^{n}\Psi(e_i)x_{ni}'H_n\delta(n)-\frac{\lambda}{2}\|\delta(n)\|^2\xrightarrow{P.}0. \tag{5.43}$$

由(5.3)、(5.4)、(5.42)、(5.43)及 $H_nH_n'+K_nK_n'=I_p$，

$$\begin{aligned}2\lambda(\tilde{M}_n-\hat{M}_n)&=\left\|\sum_{i=1}^{n}H_n'x_{ni}\Psi(e_i)\right\|^2+\lambda^2\|\delta(n)\|^2\\&\quad+2\lambda\delta(n)'\sum_{i=1}^{n}H_n'x_{ni}\Psi(e_i)+o_p(1)\\&=\left\|\sum_{i=1}^{n}H_n'x_{ni}\Psi(e_i)+\lambda\delta(n)\right\|^2+o_p(1).\end{aligned} \tag{5.44}$$

由(5.5)式、引理 5.2、(5.33)式及(5.30)式，

$$\begin{aligned}\lambda^2W_n&=\lambda^2\|H_n'\hat{\beta}(n)\|^2\\&=\left\|H_n'(\lambda K_n\gamma(n)+\lambda H_n\delta(n)+\sum_{i=1}^{n}\Psi(e_i)x_{ni})\right\|^2+o_p(1)\\&=\left\|\sum_{i=1}^{n}H_n'x_{ni}\Psi(e_i)+\lambda\delta(n)\right\|^2+o_p(1).\end{aligned} \tag{5.45}$$

由引理 5.1 及 5.2，有

$$\sum_{i=1}^{n}\Psi(Y_{ni}-x_{ni}'K_n\hat{\gamma}(n))x_{ni}-\sum_{i=1}^{n}\Psi(e_i)x_{ni}+\lambda K_n(\hat{\gamma}(n)-\gamma(n)) -\lambda H_n\delta(n)\xrightarrow{P.}0,$$

再由引理 5.2 及 $K_nK_n'+H_nH_n'=I_p$，有

$$\sum_{i=1}^{n}\Psi(Y_{ni}-x_{ni}'K_n\hat{\gamma}(n))x_{ni}-H_nH_n'\sum_{i=1}^{n}\Psi(e_i)x_{ni}-\lambda H_n\delta(n)\xrightarrow{P.}0. \tag{5.46}$$

由(5.6)、(5.10)、(5.46)及(5.30),

$$R_n = \left\| \sum_{i=1}^{n} x_{ni} \Psi(Y_{ni} - x'_{ni} K_n \hat{\gamma}(n)) \right\|^2$$

$$= \left\| \sum_{i=1}^{n} H'_n x_{ni} \Psi(e_i) + \lambda \delta(n) \right\|^2 + o_p(1). \qquad (5.47)$$

再由(5.7)、(5.10)、(5.46)及(5.30),

$$R_n^* = \left\| H'_n \sum_{i=1}^{n} x_{ni} \Psi(Y_{ni} - x'_{ni} K_n \hat{\gamma}(n)) \right\|^2$$

$$= \left\| \sum_{i=1}^{n} H'_n x_{ni} \Psi(e_i) + \lambda \delta(n) \right\|^2 + o_p(1). \qquad (5.48)$$

综合(5.44)、(5.45)、(5.47)及(5.48),即得定理5.1. ▌

这个定理的基本结果,1991年为Zhao和Chen[68]在相近的条件下证得.此处的简单证法,参考了Bai、Rao和Zhao(见文献[14])关于标准多重线性模型的有关检验统计量的渐近理论.

下面我们来证明定理5.2:

证 考虑模型(5.16),不失一般性,可设$\beta(n)=0$.

先证(5.27)式.任取$C>0$,记

$$V_n = E|\Psi(e_1 + Cd_n^{1/2}) - \Psi(e_1 - Cd_n^{1/2})|^2.$$

由Ψ的单调性,$d_n = \max_{1 \leqslant i \leqslant n} \|x_{ni}\|^2 \to 0$及A4.4,当$n$充分大时,

$$EI(\|\hat{\beta}(n)\| \leqslant C) \left| \hat{\sigma}_n^2 - n^{-1} \sum_{i=1}^{n} \Psi^2(e_i) \right|$$

$$\leqslant V_n + 2E|\Psi(e_1)[\Psi(e_1 + Cd_n^{1/2}) - \Psi(e_1 - Cd_n^{1/2})]|$$

$$\leqslant V_n + 2\sigma V_n^{1/2} \to 0. \qquad (5.49)$$

由引理5.2,有$\|\hat{\beta}(n)\| = O_p(1)$.由此及(5.49)式即得

$$\hat{\sigma}_n^2 = n^{-1} \sum_{i=1}^{n} \Psi^2(e_i) + o_p(1) \xrightarrow{P.} E\Psi^2(e_1) = \sigma^2, \text{当 } n \to \infty.$$

(5.27)式得证.

下证(5.28)式:由(5.26),

$$\hat{\lambda}_n = (2nh)^{-1} \sum_{i=1}^{n} \{\Psi(e_i - x'_{ni}\hat{\beta}(n) + h) - \Psi(e_i - x'_{ni}\hat{\beta}(n) - h)\}.$$

(5.50)

首先证明:若 $h\to 0$,$\liminf_{n\to\infty} nh^2>0$,则

$$(2nh)^{-1}\sum_{i=1}^{n}[\Psi(e_i+h)-\Psi(e_i-h)]\xrightarrow{P.}\lambda. \tag{5.51}$$

事实上,由 A4.4,有

$$\mathrm{Var}\left\{(2nh)^{-1}\sum_{i=1}^{n}[\Psi(e_i+h)-\Psi(e_i-h)]\right\}$$
$$\leqslant (4nh^2)^{-1}E[\Psi(e_1+h)-\Psi(e_1-h)]^2\to 0,\text{当 } n\to\infty.$$

因为$\lim_{n\to\infty}[G(h)-G(-h)]/(2h)=\lambda$,故

$$(2nh)^{-1}\sum_{i=1}^{n}[\Psi(e_i+h)-\Psi(e_i-h)]$$
$$=[G(h)-G(-h)]/(2h)+o_p(1)\xrightarrow{P.}\lambda,\text{当 } n\to\infty.$$

因而(5.51)式得证.

因为 $d_n^{1/2}/h_n\to 0$, $nh_n\to\infty$,我们可取 $\varepsilon_n>0$, $\varepsilon_n\to 0$,使

$$d_n^{1/2}/(\varepsilon_n h_n)\to 0,\ \text{且 } n\varepsilon_n h_n\to\infty. \tag{5.52}$$

由引理 5.2,

$$P(\|\hat{\beta}(n)\|>\varepsilon_n h_n/d_n^{1/2})\to 0. \tag{5.53}$$

记

$$g_n=I(\|\hat{\beta}(n)\|\leqslant\varepsilon_n h_n/d_n^{1/2})(2nh)^{-1}\sum_{i=1}^{n}\{\Psi(e_i-x'_{ni}\hat{\beta}(n)+h)$$
$$-\Psi(e_i-x'_{ni}\hat{\beta}(n)-h)\}. \tag{5.54}$$

由 Ψ 的单调性及(5.51)式,

$$g_n\leqslant(\varepsilon_n+1)[2n(\varepsilon_n+1)h]^{-1}$$
$$\times\sum_{i=1}^{n}[\Psi(e_i+(\varepsilon_n+1)h)-\Psi(e_i-(\varepsilon_n+1)h)]\xrightarrow{P.}\lambda. \tag{5.55}$$

同样地,且考虑到(5.53),有

$$g_n\geqslant I(\|\hat{\beta}(n)\|\leqslant\varepsilon_n h/d_n^{1/2})(2nh)^{-1}\sum_{i=1}^{n}[\Psi(e_i+(1-\varepsilon_n)h)$$
$$-\Psi(e_i-(1-\varepsilon_n)h)]\xrightarrow{P.}\lambda. \tag{5.56}$$

由(5.50)式、(5.53)式～(5.56)式即得(5.28)式. 定理 5.2 证毕. ▌

§5.2 多重线性模型中的 M 检验

在本节中,我们考虑形如(5.14)的多重线性模型,即

$$Y_i = B'x_i + \varepsilon_i \quad (i = 1,\cdots,n),$$

其中 $Y_i:p\times 1,\varepsilon_i:p\times 1,x_i:m\times 1,B:m\times p$,它们的意义在本章开头已经交代. 我们要在 M 检验的机制下考虑如下的线性假设:

$$H_0:H'B = C_0 \leftrightarrow H_1:H'B \neq C_0, \tag{5.57}$$

其中 H 为已知的 $m\times q$ 矩阵,秩为 q,C_0 为已知的 $q\times p$ 矩阵.

设 $\rho(u)$ 为 p 维变量 u 的凸函数,$\partial\rho(u)$ 为 ρ 在 u 的次微分,由附录中的 4)①,它是一个非空有界闭凸集. 我们选取一个 $\Psi(u)\in\partial\rho(u)$,即 $\Psi(u)$ 为 ρ 在 u 的一个次梯度. Ψ 作为 u 的函数是一个 p 维向量值函数. 记

$D_1=\{u\in R^p:\rho$ 在 u 可微$\}$;

$D_2=\{u\in R^p:\rho$ 在 u 有唯一的次梯度$\}$;

$D_3=\{u\in R^p:\Psi$ 在 u 连续$\}$.

由附录中的 7),$D_1=D_2$. 由附录的 5),Ψ 在任何 $u\in D_2$ 处连续,即 $D_2\subset D_3$. 下证,若 Ψ 在 $u=(u_1,\cdots,u_p)'$ 连续,则 $\partial\rho(u)$ 必为单点集. 事实上,若 $\partial\rho(u)$ 不是单点集,记 $v_1=(1,0,\cdots,0)',\cdots,v_p=(0,\cdots,0,1)'$ 为 R^p 中的一组标准正交基,则由 $\partial\rho(u)$ 为有界闭凸集,则必有 $j\quad(1\leqslant j\leqslant p)$,使 $B_j(u)=\{u_j^*=\langle u^*,v_j\rangle:u^*\in\partial\rho(u)\}$ 为一非退化的闭区间. 当局限于过 u 沿着 v_j 方向的直线上时,即若令 $\rho_1(\lambda)=\rho(u+\lambda v_j),\lambda\in(-\infty,\infty)$,则 $B_j(u)\subset\partial\rho_1(0)$,因而由附录中关于直线上凸函数的性质 11),$\Psi_1(\lambda)\triangleq\langle\Psi(u+\lambda v_j),v_j\rangle$ 在 $\lambda=0$ 处不连续,因而 Ψ 在 u 处不连续,此与假定矛盾. 由此可知,我们有 $D_1=D_2=D_3$.

---

① 本章文中提到的附录均指第 4 章的附录.

由附录的 8)，D_3 的余集 D_3^c 是一个 F_σ 型的集合，其 Lebesgue 测度为 0. 通过上面的分析可知道，对于任意选取的 Ψ，其不连续点所成的集合 D_3^c 是一样的.

我们假定 Ψ 是可测的，并作出以下的假定：

A5.1 $\varepsilon_1, \varepsilon_2, \cdots$ 独立同分布，其共同分布 F 满足 $F(D_3^c)=0$（这保证了下面出现的 Ψ 的某些泛函有唯一的值，不依赖于 Ψ 的不同选取）.

A5.2 当 $|u| \to 0$ 时，$E\Psi(\varepsilon_1+u)=\Lambda u+o(|u|)$.
式中 $\Lambda>0$ 为 $p\times p$ 常数矩阵.

A5.3 $E\Psi(\varepsilon_1)\Psi(\varepsilon_1)'=\Gamma>0$，且当 $|u|$ 充分小时，$E\|\Psi(\varepsilon_1+u)-\Psi(\varepsilon_1)\|^2$ 是有限的，它作为 u 的函数在 $u=0$ 连续.

A5.4 $S_n=\sum_{i=1}^{n} x_i x_i' > 0$，且

$$d_n=\max_{1\leqslant i\leqslant n} x_i' S_n^{-1} x_i \to 0, \text{当 } n\to\infty.$$

现分别用 $\hat{B}_n$ 和 $\widetilde{B}_n$ 记模型(5.14)中参数矩阵 B 在无限制和在

$$H'B_n = C_0 \tag{5.58}$$

限制之下的 M 估计，即 $\hat{B}_n$ 和 $\widetilde{B}_n$ 分别在无限制和在(5.58)式的限制之下，使 $\sum_{i=1}^{n} \rho(Y_i - B'x_i)$ 取得最小值. 为了检验假设(5.57)，我们提出两种检验统计量：一种是基于 $\mathrm{tr}(W_n\hat{\Lambda}_n\hat{\Gamma}_n^{-1}\hat{\Lambda}_n)$，其中

$$W_n = (H'\hat{B}_n - C_0)'(H'S_n^{-1}H)^{-1}(H'\hat{B}_n - C_0) \tag{5.59}$$

为 Wald 检验统计量，$(\hat{\Lambda}_n, \hat{\Gamma}_n)$ 为 (Λ, Γ) 的相合估计，Λ 和 Γ 分别是在 A5.2 及 A5.3 中定义的矩阵参数，稍后我们将给出这些多余参数的估计. 另一种检验是基于 $\mathrm{tr}(R_n\hat{\Gamma}_n^{-1})$，其中

$$R_n = \xi(\widetilde{B}_n)'S_n^{-1}\xi(\widetilde{B}_n) \tag{5.60}$$

为 Rao 的计分检验统计量(Rao's score type statistic 参看文献[53])，而

$$\xi(\widetilde{B}_n) = \sum_{i=1}^{n} x_i \Psi(Y_i - \widetilde{B}_n' x_i)' \tag{5.61}$$

为一 $m\times p$ 矩阵.

在给出 W_n 和 R_n 的极限分布之前,我们先简要地介绍一下 Wishart 分布(参看文献[54]中 p. 534). 设 $u_1,\cdots,u_k$ 为相互独立的 p 维正态变量,

$$u_i\sim N(\mu_i,\Sigma)\quad(i=1,\cdots,k),$$

其中 Σ 为 $p\times p$ 正定矩阵. 记 $U'=(u_1,\cdots,u_k)$, $M'=(\mu_1,\cdots,\mu_k)$, U' 与 M' 均为 $p\times k$ 矩阵,则称

$$U'U=\sum_{r=1}^{k}u_ru_r'$$

的分布是自由度为 k,非中心参数为 $M'M=\sum_{r=1}^{k}\mu_r\mu_r'$ 的非中心 Wishart 分布,记为

$$U'U\sim W_p(k,\Sigma,M'M).$$

若 $M'M=0$,则相应的分布称为自由度为 k 的中心 Wishart 分布,记为 $U'U\sim W_p(k,\Sigma)$.

若 $U'U\sim W_p(k,\Sigma,M'M)$,则 $\mathrm{tr}(U'U\Sigma^{-1})$ 服从自由度为 pk,非中心参数为 δ 的非中心 χ^2 分布 $\chi^2_{pk,\delta}$. 其中

$$\delta=\mathrm{tr}\,(M'M\Sigma^{-1}).$$

若 $M'M=0$,则 $\mathrm{tr}(U'U\Sigma^{-1})$ 服从自由度为 pk 的中心 χ^2 分布 χ^2_{pk}.

为了方便,令

$$x_{ni}=S_n^{-1/2}x_i\quad(i=1,\cdots,n);$$

$$H_n=S_n^{-1/2}H(H'S_n^{-1}H)^{-1/2};m\times q,\tag{5.62}$$

则

$$\sum_{i=1}^{n}x_{ni}x_{ni}'=I_m,\ \ H_n'H_n=I_q.\tag{5.63}$$

记

$$U_n'=\Lambda^{-1}\sum_{i=1}^{n}\Psi(\varepsilon_i)x_{ni}'H_n=(u_{1n},\cdots,u_{qn}):p\times q,\tag{5.64}$$

$$V_n'=\sum_{i=1}^{n}\Psi(\varepsilon_i)x_{ni}'H_n=(v_{1n},\cdots,v_{qn}):p\times q.\tag{5.65}$$

在本节中，我们还要考虑如下的局部对立假设：

$$H_{2,n}: H'(B - B_0) = H'\Delta_n, \tag{5.66}$$

其中 B_0 及 Δ_n 均为 $m \times p$ 矩阵，使得

$$H'B_0 = C_0, \text{而} \ \| S_n^{1/2}\Delta_n \| = O(1), \text{当} \ n \to \infty. \tag{5.67}$$

记

$$\Theta_n = H_n' S_n^{1/2} \Delta_n = (H'S_n^{-1}H)^{-1/2} H'\Delta_n : p \times q. \tag{5.68}$$

容易看出，$u_{1n}, \cdots, u_{qn}$ 是渐近独立的，并且以 p 维正态 $N(0, \Lambda^{-1}\Gamma\Lambda^{-1})$ 为其共同的极限分布，故 $U_n'U_n$ 的极限分布是自由度为 q 的中心 Wishart 分布 $W_p(q, \Lambda^{-1}\Gamma\Lambda^{-1})$. 类似地，$v_{1n}, \cdots, v_{qn}$ 也是渐近独立的，并且以 p 维正态 $N(0, \Gamma)$ 为其共同的极限分布，故 $V_n'V_n$ 的极限分布是自由度为 q 的 Wishart 分布 $W_p(q, \Gamma)$. 我们有如下定理：

定理 5.3　假定在模型(5.14)中，A5.1～A5.4 以及(5.66)、(5.67)都满足. 则当 $n \to \infty$ 时，

$$W_n = (U_n + \Theta_n)'(U_n + \Theta_n) + o_p(1). \tag{5.69}$$

特别是，若原假设 H_0 成立或当 $n \to \infty$ 时，

$$\| S_n^{1/2}\Delta_n \| \to 0,$$

则 W_n 的渐近分布为中心的 Wishart 分布 $W_p(q, \Lambda^{-1}\Gamma\Lambda^{-1})$. 若在局部对立假设之下，当 $n \to \infty$ 时，

$$\Theta_n = (H'S_n^{-1}H)^{-1/2}H'\Delta_n \text{ 有极限 } \Theta \neq 0,$$

则 W_n 的极限分布为非中心 Wishart 分布 $W_p(q, \Lambda^{-1}\Gamma\Lambda^{-1}, \Theta'\Theta)$.

定理 5.4　假定在模型(5.14)之下，A5.1～A5.4 及(5.66)、(5.67)都满足，则当 $n \to \infty$ 时，

$$R_n = (V_n + \Theta_n\Lambda)'(V_n + \Theta_n\Lambda) + o_p(1). \tag{5.70}$$

特别是，若 H_0 成立，或当 $n \to \infty$ 时，$\| S_n^{1/2}\Delta_n \| \to 0$，则 R_n 的极限分布为中心的 Wishart 分布 $W_p(q, \Gamma)$. 若当 $n \to \infty$ 时，Θ_n 有极限 $\Theta \neq 0$，则 R_n 的渐近分布为非中心 Wishart 分布 $W_p(q, \Gamma, \Lambda\Theta'\Theta\Lambda)$.

定理 5.3 和 5.4 的证明，本质上与定理 5.1 的证明类似，因而在此从略. 有兴趣的读者可以参看文献[14].

这里需要指出，当由 W_n 构造检验统计量 $\mathrm{tr}(W_n\hat{\Lambda}_n\hat{\Gamma}_n^{-1}\hat{\Lambda}_n)$ 时，我们需要估计两个多余矩阵参数 Γ 和 Λ；而由 R_n 构造检验统计量 $\mathrm{tr}(R_n\hat{\Gamma}_n^{-1})$ 时，我们仅需要估计一个多余矩阵参数 Γ. 由于在局部对立假设之下，$\mathrm{tr}(W_n\hat{\Lambda}_n\hat{\Gamma}_n^{-1}\hat{\Lambda}_n)$ 和 $\mathrm{tr}(R_n\hat{\Gamma}_n^{-1})$ 都渐近等价于自由度为 pq，非中心参数为

$$\delta_n = \mathrm{tr}(\Theta_n'\Theta_n\Lambda\Gamma^{-1}\Lambda) \tag{5.71}$$

的 χ^2 变量，故在局部对立假设之下，由此得出的两个检验有渐近相同的效. 以上的陈述只是对大样本成立. 这两种检验的优缺点，还需要通过小样本情形的研究来加以比较.

下面我们讨论多余参数矩阵 Γ 和 Λ 的估计. Γ 的一个自然的估计是：

$$\hat{\Gamma}_n = n^{-1}\sum_{i=1}^{n}\Psi(Y_i - \hat{B}_n'x_i)\Psi(Y_i - \hat{B}_n'x_i)', \tag{5.72}$$

此处 $\hat{B}_n$ 是模型(5.14)中 B 的 M 估计. 为估计 Λ，可任取一个 $p\times p$ 非奇异矩阵 $Z=(\zeta_1,\cdots,\zeta_p)$，其中 p 维向量 ζ_j 为它的第 j 列 $(1\leqslant j\leqslant n)$. 取 $h=h_n>0$，使

$$h_n/d_n^{1/2} \to \infty,\ h_n \to 0,\ \liminf_{n\to\infty} nh_n^2 > 0. \tag{5.73}$$

定义

$$\eta_k = (2nh)^{-1}\sum_{i=1}^{n}\{\Psi(Y_i - \hat{B}_n'x_i + h\zeta_k) - \Psi(Y_i - \hat{B}_n'x_i - h\zeta_k)\}$$
$$(k = 1,\cdots,p), \tag{5.74}$$

$$A_n = (\eta_1,\cdots,\eta_p)Z^{-1}, \tag{5.75}$$

并用 p 阶对称阵

$$\hat{\Lambda}_n = \frac{1}{2}(A_n + A_n') \tag{5.76}$$

作为正定阵 Λ 的一个估计.

我们有下面的定理：

定理 5.5 设在模型(5.14)之下，A5.1～A5.4 成立，则

$$\hat{\Gamma}_n \xrightarrow{P.} \Gamma,\text{当 } n\to\infty. \tag{5.77}$$

进一步，若(5.73)式成立，则

$$\hat{\Lambda}_n \xrightarrow{P.} \Lambda, \text{当 } n \to \infty. \tag{5.78}$$

因为此定理的证明方法比较独特，因此，我们按照文献[14]中的证法给出它的证明. 记

$$x_{ni} = S_n^{-1/2} x_i \quad (1 \leqslant i \leqslant n); \quad B(n) = S_n^{1/2} B, \quad \hat{B}(n) = S_n^{1/2} \hat{B}_n.$$

在模型(5.14)的规范形式

$$Y_i = B(n)' x_{ni} + \varepsilon_i \quad (i = 1, \cdots, n) \tag{5.79}$$

之下，$\hat{B}(n)$为 $B(n)$的 M 估计. 用定理 4.1 中类似的方法可证明

$$\hat{B}(n) - B(n) = \sum_{i=1}^{n} x_{ni} \Psi(\varepsilon_i)' \Lambda^{-1} + o_p(1). \tag{5.80}$$

有了这些，我们就可以证明定理 5.5 了.

证 令 $\quad u = (u_1, \cdots, u_p)', \ v = (v_1, \cdots, v_p)',$

$$\Psi(u) = (\Psi_1(u), \cdots, \Psi_p(u))'.$$

记 $\quad T = \{\tau = (\tau_1, \cdots, \tau_p)' : \tau_1, \cdots, \tau_p = \pm 1\}.$

首先我们证明，若$|v_k| \leqslant b/2$ 对某个 $b > 0$ 和 $k = 1, \cdots, p$ 成立，则存在一个常数 $C > 0$，使

$$C \sum_{\tau \in T} \tau' [\Psi(u - b\tau) - \Psi(u)] \leqslant \Psi_1(u + v) - \Psi_1(u) \leqslant C \sum_{\tau \in T} \tau' [\Psi(u + b\tau) - \Psi(u)], \tag{5.81}$$

类似的不等式对 $\Psi_k(u+v) - \Psi_k(u)$ $(k = 2, \cdots, p)$也成立.

我们指出，对任意的 $\tau \in T$，

$$\tau' [\Psi(u + b\tau) - \Psi(u)] \geqslant 0,$$

而

$$\tau' [\Psi(u - b\tau) - \Psi(u)] \leqslant 0.$$

事实上，由凸函数的次微分的循环单调性(见第 4 章附录中 6))，对$\forall \tau \in T$，有

$$v' \Psi(u) + (b\tau - v)' \Psi(u + v) - b\tau' \Psi(u + b\tau) \leqslant 0,$$

它蕴涵了

$$(b\tau - v)' [\Psi(u + v) - \Psi(u)] \leqslant b\tau' [\Psi(u + b\tau) - \Psi(u)] \tag{5.82}$$

以及

$$(b\tau+v)'[\Psi(u+v)-\Psi(u)]\geqslant b\tau'[\Psi(u-b\tau)-\Psi(u)]. \tag{5.83}$$

为简单起见，置 $\tilde{\tau}=(\tau_1,\cdots,\tau_{p-1})'$，$\tilde{u}=(u_1,\cdots,u_{p-1})'$，$\tilde{v}=(v_1,\cdots,v_{p-1})'$，$\tilde{\Psi}=(\Psi_1,\cdots,\Psi_{p-1})'$，有时我们也将 $\Psi(u)$ 写成 $\Psi(u')$. 在(5.82)中分别取 $\tau_p=1$、-1，可得

$$\begin{aligned}&(b\tilde{\tau}'-\tilde{v}')(\tilde{\Psi}(u+v)-\tilde{\Psi}(u))\\&\quad+(b-v_p)(\Psi_p(u+v)-\Psi_p(u))\\&\leqslant b(\tilde{\tau}',1)[\Psi(\tilde{u}'+b\tilde{\tau}',u_p+b)-\Psi(u)]\end{aligned} \tag{5.84}$$

及

$$\begin{aligned}&(b\tilde{\tau}'-\tilde{v}')(\tilde{\Psi}(u+v)-\tilde{\Psi}(u))\\&\quad-(b+v_p)(\Psi_p(u+v)-\Psi_p(u))\\&\leqslant b(\tilde{\tau}',-1)[\Psi(\tilde{u}'+b\tilde{\tau}',u_p-b)-\Psi(u)].\end{aligned} \tag{5.85}$$

将(5.85)式的两端同乘以 $(b-v_p)/(b+v_p)$，并分别加到(5.84)式的两端，即可消去 $\Psi_p(u+v)-\Psi_p(u)$，得到

$$\begin{aligned}&(2b/(b+v_p))(b\tilde{\tau}'-\tilde{v}')(\tilde{\Psi}(u+v)-\tilde{\Psi}(u))\\&\quad\leqslant b(\tilde{\tau}',1)[\Psi(u'+b(\tilde{\tau}',1))-\Psi(u)]\\&\qquad+b(\tilde{\tau}',-1)[\Psi(u'+b(\tilde{\tau}',-1))-\Psi(u)]\frac{b-v_p}{b+v_p}.\end{aligned} \tag{5.86}$$

用这种逐步消去法即可得到(5.81)的第二个不等式. (5.81)的第一个不等式可由(5.83)式类似地得到.

为证明定理 5.5，不失一般性，可设在模型(5.14)中真参数矩阵 $B=0$，即模型(5.79)中 $B(n)=0$. 由(5.80)式，

$$\|\hat{B}(n)\|=O_p(1). \tag{5.87}$$

由 $d_n\to0$，当 $n\to\infty$ 时，

$$P(\|\hat{B}(n)\|\geqslant d_n^{-1/4})\to0. \tag{5.88}$$

由 A5.1、A5.3 和强大数律，当 $n\to\infty$ 时，

$$\tilde{\Gamma}\equiv n^{-1}\sum_{i=1}^{n}\Psi(\varepsilon_i)\Psi(\varepsilon_i)'\to\Gamma\equiv(\gamma_{lm}),\ \text{a.s.}. \tag{5.89}$$

记 $\hat{\Gamma}_n=(\hat{\gamma}_{lm}),\widetilde{\Gamma}=(\widetilde{\gamma}_{lm})$，有

$$|\hat{\gamma}_{lm}-\widetilde{\gamma}_{lm}|^2=\Big|n^{-1}\sum_{i=1}^{n}[\Psi_l(\varepsilon_i-\hat{B}(n)'x_{ni})\Psi_m(\varepsilon_i-\hat{B}(n)'x_{ni})-\Psi_l(\varepsilon_i)\Psi_m(\varepsilon_i)]\Big|^2$$

$$\leqslant n^{-1}\sum_{i=1}^{n}[\Psi_l(\varepsilon_i-\hat{B}(n)'x_{ni})-\Psi_l(\varepsilon_i)]^2\times n^{-1}\sum_{i=1}^{n}\Psi_m^2(\varepsilon_i-\hat{B}(n)'x_{ni})+n^{-1}\sum_{i=1}^{n}\Psi_l^2(\varepsilon_i)\times n^{-1}\sum_{i=1}^{n}[\Psi_m(\varepsilon_i-\hat{B}(n)'x_{ni})-\Psi_m(\varepsilon_i)]^2. \tag{5.90}$$

在事件 $\{\|\hat{B}(n)\|<d_n^{-1/4}\}$ 发生时，$\|\hat{B}(n)'x_{ni}\|<d_n^{1/4}$ 对任意的 $i\leqslant n$. 由(5.81)式，存在一个正的常数 C，使

$$n^{-1}\sum_{i=1}^{n}[\Psi_l(\varepsilon_i-\hat{B}(n)'x_{ni})-\Psi_l(\varepsilon_i)]^2 I(\|\hat{B}(n)\|<d_n^{-1/4})\leqslant C\cdot\max\Big\{n^{-1}\sum_{i=1}^{n}\|\Psi(\varepsilon_i+2d_n^{1/4}\tau)-\Psi(\varepsilon_i)\|^2:\tau\in T\Big\}. \tag{5.91}$$

由 A5.3、A5.4，对固定的 $\tau\in T$，当 $n\to\infty$ 时，

$$E\Big\{n^{-1}\sum_{i=1}^{n}\|\Psi(\varepsilon_i+2d_n^{1/4}\tau)-\Psi(\varepsilon_i)\|^2\Big\}=E\|\Psi(\varepsilon_1+2d_n^{1/4}\tau)-\Psi(\varepsilon_1)\|^2\to 0. \tag{5.92}$$

故由(5.88)～(5.92)，注意到 $\#(T)=2^p$，有

$$\lim_{n\to\infty}|\hat{\gamma}_{lm}-\widetilde{\gamma}_{lm}|\xrightarrow{P.}0\quad(l,m=1,\cdots,p). \tag{5.93}$$

由(5.89)及(5.93)即得(5.77)式.

今证(5.78)式：为此，我们证明对任何常数 $C>0$，

$$(nh)^{-1}\sum_{i=1}^{n}[\Psi_l(\varepsilon_i-\hat{B}(n)'x_{ni}+h\zeta_k)-\Psi_l(\varepsilon_i+h\zeta_k)]\times I(\|\hat{B}(n)\|\leqslant C)\xrightarrow{P.}0\quad(l=1,\cdots,p). \tag{5.94}$$

由(5.81)式，为证(5.94)，只需证明对每个固定的 $\tau\in T$，有

$$V_n \equiv (nh)^{-1}\sum_{i=1}^{n}\tau'[\Psi(\varepsilon_i + 2Cd_n^{1/2}\tau + h\zeta_k)$$
$$- \Psi(\varepsilon_i + h\zeta_k)] \xrightarrow{P.} 0. \tag{5.95}$$

由 A5.1、A5.3、A5.4 及(5.73)式，

$$\mathrm{Var}(V_n) \leqslant (nh^2)^{-1}E\{\tau'[\Psi(\varepsilon_1 + 2Cd_n^{1/2}\tau + h\zeta_k)$$
$$- \Psi(\varepsilon_1 + h\zeta_k)]\}^2$$
$$\leqslant (nh^2)^{-1}\|\tau\|^2 E\|\Psi(\varepsilon_1 + 2Cd_n^{1/2}\tau + h\zeta_k)$$
$$- \Psi(\varepsilon_1 + h\zeta_k)\|^2 \to 0. \tag{5.96}$$

另一方面，由 A5.2、A5.4 及(5.73)式，

$$EV_n = h^{-1}\tau' E[\Psi(\varepsilon_1 + 2Cd_n^{1/2}\tau + h\zeta_k)$$
$$- \Psi(\varepsilon_1 + h\zeta_k)] \to 0. \tag{5.97}$$

由(5.96)及(5.97)即得(5.95)及(5.94). 注意 $\|\hat{B}(n)\| = O_p(1)$，得

$$(nh)^{-1}\sum_{i=1}^{n}[\Psi(\varepsilon_i - \hat{B}(n)'x_{ni} + h\zeta_k) - \Psi(\varepsilon_i + h\zeta_k)] \xrightarrow{P.} 0. \tag{5.98}$$

同样有

$$(nh)^{-1}\sum_{i=1}^{n}[\Psi(\varepsilon_i - \hat{B}(n)'x_{ni} - h\zeta_k) - \Psi(\varepsilon_i - h\zeta_k)] \xrightarrow{P.} 0. \tag{5.99}$$

由 A5.1、A5.3 及(5.73)式，对 $m=1,\cdots,p$，当 $n\to\infty$ 时，

$$\mathrm{Var}\{(nh)^{-1}\sum_{i=1}^{n}[\Psi_m(\varepsilon_i \pm h\zeta_k) - \Psi_m(\varepsilon_i)]\}$$
$$\leqslant (nh^2)^{-1}E[\Psi_m(\varepsilon_1 \pm h\zeta_k) - \Psi_m(\varepsilon_1)]^2 \to 0. \tag{5.100}$$

另一方面，由 A5.2 及 $h_n\to 0$，有

$$(2nh)^{-1}\sum_{i=1}^{n}E[\Psi(\varepsilon_i + h\zeta_k) - \Psi(\varepsilon_i - h\zeta_k)]$$
$$= (2h)^{-1}E[\Psi(\varepsilon_1 + h\zeta_k) - \Psi(\varepsilon_1 - h\zeta_k)]$$
$$\to \Lambda\zeta_k \quad (k=1,\cdots,p). \tag{5.101}$$

由(5.100)及(5.101),对 $k=1,\cdots,p$,

$$(2nh)^{-1}\sum_{i=1}^{n}[\Psi(\epsilon_i+h\zeta_k)-\Psi(\epsilon_i-h\zeta_k)]\xrightarrow{P.}\Lambda\zeta_k. \tag{5.102}$$

由(5.74)、(5.98)、(5.99)及(5.102)式,并注意 $\hat{B}_n'x_i=\hat{B}(n)'x_{ni}$,可得

$$\eta_k\xrightarrow{P.}\Lambda\zeta_k\quad(k=1,\cdots,p),\text{当 } n\to\infty. \tag{5.103}$$

由(5.75)及(5.103),

$$A_n\xrightarrow{P.}\Lambda(\zeta_1,\cdots,\zeta_p)Z^{-1}=\Lambda.$$

故也有 $\hat{\Lambda}_n\xrightarrow{P.}\Lambda$,从而(5.78)式得证. 这样就完成了定理 5.5 的证明. ▌

§5.3 历 史 小 记

将 M 方法用于线性模型(5.1)之下线性假设(5.2)的检验,较早的工作可以举出 Schrader 和 Hettmansperger[60]及 Sen[61]的工作. 1980 年,Schrader 和 Hettmansperger[60]在 Huber 的文献[34]中的三个条件(参看 §4.3)之下研究了 $\tilde{M}_n-\hat{M}_n$(参看(5.3)及(5.4)式)在 H_0 及一列局部对立假设之下的渐近分布,文中对凸函数 ρ 的光滑性提出了较高的要求,需要有足够高阶的有界导函数. 1982 年,Sen[61]研究了 Rao 的计分检验准则的一种变化形式,即(5.7)式中定义的 R_n^*. 为了建立 R_n^* 在 H_0 之下的渐近的 χ^2 分布,他对 Ψ、F 和 $\{x_i\}$ 加上了较强的限制条件,其中包括要求随机误差 e_1 的分布 F 具有绝对连续的密度函数 f 和有限的 Fisher 信息量:

$$I(f)=\int_{-\infty}^{\infty}(f'(x)/f(x))^2\mathrm{d}F(x).$$

至于加在 $\{x_i\}$ 上的、Jurečkova[39]引进的复杂条件,在 Singer 和 Sen 于 1985 年研究多重线性模型的 M 检验的文章[62]中已经去掉

了.

在 $\rho(u)=|u|$ 的特殊情形,即导致了最小绝对偏差(LAD)的检验准则,在这方面的参考文献有: Koenker 和 Bassett 1982 年的论述[42]、Koenker 1987 年的论文[41]、Mckean 和 Schrader 1987 年的论文[47]、Bai、Rao 和 Yin 1990 年的论述[13]等. 用 $\tilde{\beta}_n$ 和 $\hat{\beta}_n$ 分别记在有关的线性限制下和无限制时回归参数 β 的 LADE,在文献[13]中,在较弱的条件下,对 $\sum_{i=1}^{n}|Y_i-x_i'\tilde{\beta}_n|-\sum_{i=1}^{n}|Y_i-x_i'\hat{\beta}_n|$ 的渐近分布给出了严格的推导.

1991 年,Zhao 和 Chen 在文献[68]中在模型(5.1)之下,研究了线性假设(5.2)的两种检验准则 $\tilde{M}_n-\hat{M}_n$ 及 R_n^*,在一列局部对立假设 $H_{2,n}$ 下,在与 A4.1~A4.5 相近的条件下建立了它们的渐近表达式和渐近分布,并且给出了多余参数的相合估计.

考虑形如

$$Y_i=X_i'\beta+\varepsilon_i\quad(i=1,\cdots,n)\tag{5.104}$$

的多重线性模型,其中 Y_i 为 p 维观测向量,X_i 为已知的 $m\times p$ 矩阵,β 为未知的 m 维回归参数,$\{\varepsilon_i\}$ 为一列 iid. p 维随机误差. 设 H 是秩为 q 的 $m\times q$ 矩阵,β_0 为已知的 m 维向量. 为检验线性假设 H_0:$H'(\beta-\beta_0)=0$,1990 年 Bai、Chen、Miao 和 Rao[11]研究了在"最小距离"意义下的有关检验准则. 用 $\tilde{\beta}_n$ 和 $\hat{\beta}_n$ 分别记在 H_0 限制下和没有限制时 β 的最小距离估计,即

$$\sum_{i=1}^{n}\|Y_i-X_i'\tilde{\beta}_n\|=\inf_{H_0}\sum_{i=1}^{n}\|Y_i-X_i'\beta\|,$$

$$\sum_{i=1}^{n}\|Y_i-X_i'\hat{\beta}_n\|=\inf_{\beta\in R^p}\sum_{i=1}^{n}\|Y_i-X_i'\beta\|.$$

他们在较弱的条件下建立了在 H_0 之下

$$\sum_{i=1}^{n}\|Y_i-X_i'\tilde{\beta}_n\|-\sum_{i=1}^{n}\|Y_i-X_i'\hat{\beta}_n\|$$

的极限的加权 χ^2 分布. 在误差分布为球对称的条件下,上述极限分布可以为自由度为 q 的中心 χ^2 分布.

1992 年,Bai、Rao 和 Wu[12]在 M 机制下研究了在多重模型

(5.104)之下 H_0 的假设检验. 用 $\tilde{\beta}_n$ 和 $\hat{\beta}_n$ 分别记 β 的 M 估计. 在很弱的条件下他们证明了:在 H_0 之下,

$$\sum_{i=1}^{n}\rho(Y_i - X_i'\tilde{\beta}_n) - \sum_{i=1}^{n}\rho(Y_i - X_i'\hat{\beta}_n)$$

渐近于一个加权的 χ^2 分布,而相应的 Wald 统计量渐近于自由度为 q 的中心 χ^2 分布.

1993 年,Bai、Rao 和 Zhao[14]研究了在标准的多重线性模型(5.14)之下检验假设 $H_0: H'B=C_0$ 的问题,在一列局部对立假设 $H_{2,n}$之下,建立了 Wald 检验统计量和 Rao 的计分检验统计量的 Wishart 渐近分布,即 § 5.2 中的定理 5.3 和定理 5.4,并且给出了多余矩阵参数的相合估计.

第 6 章

M 估计的线性表示

§6.1 引言与启发式推导

在第 4 章中证明 M 估计的渐近正态性时，我们所使用的方法的要旨，在于从 M 估计中分出一个线性独立和，即独立随机变量的线性组合，它与 M 估计之差称为剩余项，在概率上是无穷小. 这样，M 估计的极限分布与这线性独立和的极限分布是一致的. 而根据古典中心极限定理知，后者在很一般的条件下就是正态分布，这就得出了所要的结果. 当然，这个证明尽管在概念上简单清楚，但却并非容易实现. 因为找出上述线性独立和以及证明剩余项在概率上可忽略不计等，均非易事.

这个方法有相当程度的普遍性，我们所熟知的一些常用统计量，如矩估计、极大似然估计，其渐近正态性的证明，用的就是这种方法——对原点矩估计直接用古典中心极限定理，而对中心矩估计，利用它可表为若干个原点矩估计的多项式，对此多项式用 Taylor 展开，分出其线性部分(由于各原点矩估计并非独立，需要用到多维的古典中心极限定理). 对极大似然估计，则通过似然方程分出其线性部分. 在此，剩余项的处理就稍稍复杂一点了. 另外一些重要的例子，如 Hoeffding 的 U 统计量、线性秩统计量、次序统计量线性组合等(见文献[3])，其渐近正态性的证明也是用这个想法. 当然，这并不是说，此乃处理统计量渐近分布的唯一方法. 在渐近分布为非正态的情况，如有名的 Kolmogorov－Smirnov 统计

量、极值统计量等，这个方法就不适用了. 但由于正态分布是统计量的最常见的一种极限分布，而经验表明，在这种情况下此法往往可用，这使得它仍不失为最重要的一种方法.

一般地，设 $Y_1,\cdots,Y_n$ 为独立样本，$T_n=T_n(Y_1,\cdots,Y_n)$ 是一个统计量或其一个线性变换（例如，$\hat{\theta}_n=\hat{\theta}_n(Y_1,\cdots,Y_n)$ 是某参数向量 θ 的一个估计. 设 θ 为 p 维，a_n 和 A_n 分别是已知的或依赖于 θ 的 p 维向量和 p 阶非异方阵，令

$$T_n = A_n(\hat{\theta}_n - a_n).$$

按照这种提法，a_n 和 A_n 可以依赖于参数，严格说来，T_n 不必是统计量. 这一点对下文的讨论并无影响）. 所谓 T_n 的线性表示，是指把 T_n 表为

$$T_n = L_n + R_n,\ L_n = c_{n1}z_1 + \cdots + c_{nn}z_n, \tag{6.1}$$

其中 $z_1,\cdots,z_n$ 线性独立. 一般地 z_i 只依赖于 Y_i，从而其独立性由 Y_i 的独立性保证. $c_{n1},\cdots,c_{nn}$ 是常数. L_n 是 T_n 的线性部分（或线性项），而 R_n 称为*剩余项*；R_n 必须在某种意义上为无穷小（随 $n\to\infty$ 而趋于 0），不然，表达式(6.1)对于探讨 T_n 的渐近性质就没有意义了（从实用的观点看，如果 R_n 虽非无穷小，但相对于 L_n 而言甚小，则简化了的 L_n 可近似地取代 T_n 而使统计方法简化. 这一点与本章的讨论无关，表过不再提了）. 最常用的两种情况是，R_n 依概率无穷小（$o_p(1)$），及以概率 1 或几乎必然无穷小（$o(1)$，a.s.）. 对这两种情况，(6.1)式中的 L_n 分别称为 T_n 的*弱（线性）表示*和*强（线性）表示*. 对深一层的渐近性质而言，往往仅要求 R_n 为 $o_p(1)$ 或 $o(1)$，a.s. 还不够，还须它趋于 0 时有一定的速度.

总结上述，就可以把统计量的线性表示的问题一般地表述为：要找出表达式(6.1)（这等于说要定出 L_n），使在尽可能弱的条件下，证明剩余项有尽可能快的趋于 0 的速度. 不言而喻，这“尽可能”可走多远，取决于 T_n，而并非随心所欲，然而，所用的方法及分析的深度，决定了所得结果的深度.

在作了上述一般性的开场白以后，让我们回到线性模型中的 M 估计的问题. 仍设有线性模型(1.1)，(1.1)式中的 x_i 和 β_0 分别

为已知的和未知的 p 维常向量,随机误差 $e_1,e_2,\cdots$ 假定为 iid.,本章中一些结果不难推广到 e_i 仅是独立但不必为同分布的情况.这种推广只是在形式上复杂一点,推理的本质并无变化.选定函数 ρ,仍以 $\hat{\beta}_n\equiv(\hat{\beta}_{n1},\cdots,\hat{\beta}_{np})\equiv\hat{\beta}_n(Y_1,\cdots,Y_n)$ 记使用函数 ρ 所产生的 M 估计.我们先提出 $\hat{\beta}_n$ 的线性表示的形式,再给出一个启发式的(形式的,非严格的)推导.在以下几节中将给出严格的论证,并指出这一线性表示的若干应用.

记

$$S_n=\sum_{i=1}^n x_i x_i',$$

$$x_{ni}=S_n^{-1/2}x_i \quad (1\leqslant i\leqslant n),$$

又暂设 $\Psi(u)=\rho'(u)$ 当 $u\in R^1$ 处处存在,且 $\Psi'(u)$ 也处处存在,$E\Psi'(e_1)$ 存在、有限,且不为 0:$h\equiv E\Psi'(e_1)\neq 0$,则

$$S_n^{1/2}(\hat{\beta}_n-\beta_0)=h^{-1}\sum_{i=1}^n\Psi(e_i)x_{ni}+R_n. \tag{6.2}$$

这里,最初的对象是 M 估计 $\hat{\beta}_n$,$S_n^{1/2}(\hat{\beta}_n-\beta_0)$ 是 $\hat{\beta}_n$ 的一个线性变换,相当于(6.1)式中的 T_n,$\Psi(e_i)$ 是(6.1)式中的 z_i.

为导出(6.2),利用 M 估计的定义及 $\Psi=\rho'$ 处处存在的条件,得

$$\sum_{i=1}^n\Psi(Y_i-x_i'\hat{\beta}_n)x_i=0.$$

注意,$Y_i-x_i'\beta_0=e_i$,上式可改写为:

$$\sum_{i=1}^n\Psi(e_i-x_i'(\hat{\beta}_n-\beta_0))x_i'=0. \tag{6.3}$$

将 $\Psi(e_i-x_i'(\hat{\beta}_n-\beta_0))$ 在 e_i 处作 Taylor 展开,只取其线性部分,近似地有

$$\Psi(e_i-x_i'(\hat{\beta}_n-\beta_0))\approx\Psi(e_i)-\Psi'(e_i)(\hat{\beta}_n-\beta_0)'x_i. \tag{6.4}$$

上式当 n 充分大时,近似成立.其直观理由是:在很一般的条件下,$\hat{\beta}_n$ 为相合估计,故当 n 很大时,$\hat{\beta}_n-\beta_0$ 充分小,因而高次项可忽略

不计. 以(6.4)代入(6.3),得

$$\sum_{i=1}^{n}\Psi'(e_i)x_ix_i'(\hat{\beta}_n-\beta_0)\approx\sum_{i=1}^{n}\Psi(e_i)x_i. \tag{6.5}$$

由于 $h=E\Psi'(e_i)\neq 0$,在一般条件下,按大数定律,当 n 充分大时,近似地有

$$\sum_{i=1}^{n}\Psi'(e_i)x_ix_i'\approx h\sum_{i=1}^{n}x_ix_i'=hS_n. \tag{6.6}$$

以(6.6)代入(6.5),得

$$hS_n(\hat{\beta}_n-\beta_0)\approx\sum_{i=1}^{n}\Psi(e_i)x_i.$$

在上式两端左乘以 $h^{-1}S_n^{-1/2}$,即得

$$S_n^{1/2}(\hat{\beta}_n-\beta_0)\approx\sum_{i=1}^{n}\Psi(e_i)S_n^{-1/2}x_i/h=\sum_{i=1}^{n}\Psi(e_i)x_{ni}/h. \tag{6.7}$$

这样就得到了(6.2), 其中剩余项 R_n 为无穷小.

当然,这个启发式推导的目的,只是在探究如果 $\hat{\beta}_n$ 有线性表示,它该有什么样的形状,而不是严格的证明. 甚至在以下各节的严格理论中,上述推演过程中的某些假定也不必要,例如 ρ 处处可微的条件(Ψ 的定义略加修改). 我们所要做的是:在尽可能少的条件下去证明(6.2)式中的 R_n 为无穷小.

这个领域的工作最早发源于 Bahadur 在 1966 年的文章[10],其中研究了样本分位数的线性表示,对样本中位数的特例,相当于模型 $Y_i=\beta_0+e_i$ 且取 $\rho(u)=|u|$ 的情况. 记 $\hat{\beta}_n=\mathrm{med}(Y_1,\cdots,Y_n)$. 设 $e_1,e_2,\cdots$ 为 iid. ,$\mathrm{med}(e_1)=0$, e_1 的分布 F 在 0 点有导数 $f(0)>0$. 在这种条件下,很容易得出 $\hat{\beta}_n$ 的弱线性表示:

$$\sqrt{n}(\hat{\beta}_n-\beta_0)=\sum_{i=1}^{n}\mathrm{sgn}(e_i)/(2\sqrt{n}f(0))+R_n, \tag{6.8}$$

而剩余项 R_n 有数量级 $O_p(n^{-1/4})$(其证明见文献[3]中引理 1.2). 较为困难的是强表示,其形式与(6.8)相同. 在更强一些的条件下, Bahadur 证明剩余项 R_n 有数量级:

$$R_n=O(n^{-1/4}(\log n)^{1/2}(\log\log n)^{1/4}),\ \text{a.s.}. \tag{6.9}$$

(6.9)这个数量级远不如弱表示意义下的数量级 $O_p(n^{-1/4})$,但这并

非由于证法上的缺陷(详见后). 事实上可以证明,(6.9)式中的主要部分 $n^{-1/4}$ 已无可改进,即不能以 n^{-a} 代替 $n^{-1/4}$,对任何 $a>1/4$.

在 Bahadur 以后,这个方面的工作有多年没有什么进展,这并不奇怪,因为有关线性模型参数 M 估计的基本工作,到 1973 年才由 Huber 提出来,其中只涉及诸如相合性和渐近正态性这类较浅层次的渐近性质问题. 在此后若干年中,M 估计的研究大体仍局限于这种问题. 直到 1989 年,Babu[9]研究了在一般线性模型(1.1)之下 LADE 的线性表示. 1992 年,Rao 和 Zhao[57]把 Babu 的研究推广到一般凸函数的情形. 1993 年,陈希孺于文献[2]中,在限制性条件之下研究了 ρ 不必为凸函数时,M 估计的线性表示,并考虑了其若干应用. 本章的内容即是以上述这两项工作为基础的.

可以把统计量的线性表示与统计量分布的渐近展开作一比较. 二者相似之处在于:其目的都是设法用一个较简单、较易处理的对象去逼近一个复杂的、难于直接处理的量. 不同之处在于:统计量的线性表示是针对统计量本身;而统计量分布的渐近展开则是针对其分布. 另外,线性表示其逼近的阶有一定限度;而渐近展开在适当的条件下,其逼近的阶可以愈来愈高.

本章将限于讨论 x_i 为非随机的情况. 如以前曾指出过的,当 x_i 本身也是随机时,有关的结果原则上可以由 x_i 为非随机的情况推出来. 但是,如同我们在讨论弱相合的问题时曾看到的,这样做往往使所提的条件不必要地加强了. 因此,还有必要直接去研究这种情况. 对此,我们介绍读者参阅文献[48].

§6.2 ρ 为凸函数时的线性表示

取线性模型(1.1),并仍沿用前面的记号:以 $\hat\beta_n$ 记 β_0 的用 ρ 所产生的 M 估计,令 $S_n=\sum_{i=1}^n x_i x_i'$,以及

$$x_{ni}=S_n^{-1/2}x_i \quad (1\leqslant i\leqslant n),\quad d_n=\max_{1\leqslant i\leqslant n}\|x_{ni}\|^2.$$

并分别以 λ_n 和 μ_n 记 S_n 的最小和最大特征根. 对任意两串常数

$\{a_n\}$和$\{b_n\}$，若 $a_n=O(b_n)$且 $b_n=O(a_n)$，则记为 $a_n\sim b_n$. 以 Ψ_- 和 Ψ_+分别记 ρ 的左、右导数.

本节的内容是讨论当 ρ 为凸函数时，线性表示(6.2)中剩余项的阶. 如前面提到的，Rao 和 Zhao 在 1992 年的工作[57]是第一个用一般的形式研究这个问题的. 本节所用的方法就是本源于文献[57]. 但在这里增加了对一个特殊情况——即 $\lambda_n\sim\mu_n$ 情况的讨论——在这个情况下，文献[57]中得到的 R_n 的阶可以改善. $\lambda_n\sim\mu_n$ 成立的情况在应用上有重要性，因为它包含了

$$S_n/n \to V > 0 \tag{6.10}$$

的特例.

一、$\lambda_n\sim\mu_n$ 满足的情况：定理表述与假定的讨论

定理 6.1　假定有以下的条件：

A₁　$e_1,e_2,\cdots$为 iid.；

A₂　ρ 为凸函数，左、右导数为 Ψ_-、Ψ_+；

A₃　存在 Ψ，$\Psi_-\leqslant\Psi\leqslant\Psi_+$，使 $E\Psi(e_1)=0$；

A₄　当$|u|$充分小时，$G(u)\equiv E\Psi(e_1+u)$存在有限；

A₅　$0<\sigma^2\equiv E\Psi^2(e_1)<\infty$；

A₆　$E|\Psi(e_1)|^3<\infty$；

A₇　对充分小的 $h>0$，有 $\sup\limits_{|u|<\infty}|\Psi(u+h)-\Psi(u)|<\infty$；

A₈　当$|u|$充分小时，$G'(u)$存在. 记 $g=G'(0)$，有$|G'(u)-g|=O(|u|^{1/2})$，当 $u\to 0$；

A₉　$d_n=o((\log n)^{-2}(\log\log\mu_n)^{-1})$，且对某个 $\delta>2$，有 $d_n=O((\log\mu_n)^{-\delta})$；

A₁₀　$\lambda_n\sim\mu_n$.

则如在(6.2)式中以 g 取代 h，剩余项 R_n 满足

$$R_n = O(d_n^{1/4}(\log n)^{1/2}(\log\log\mu_n)^{1/4}),\ \text{a.s.}. \tag{6.11}$$

看起来定理的条件罗列很多，其实内涵并不复杂. 下面作些解释：

1）$A_1\sim A_5$ 这些条件以前在讨论相合性时曾提出过. 线性表

示的要求比相合性高得多，故其出现应属自然. A_6 要求 $\Psi(e_1)$的三阶矩存在有限，比 A_5 略强些. 基于问题的更深的层次，这也可以认为是不过分的.

2）条件 A_9 要求 $d_n\to 0$，并有一定的速度. 这种条件在前面讨论 $\hat{\beta}_n$ 的强相合性和渐近正态性时也曾碰到过. 线性表示是更深层次的问题，出现这种条件，且对 d_n 的阶要求更高，是可以理解的——我们进一步看出了 d_n 这个量在 $\hat{\beta}_n$ 的大样本理论中的作用.

A_9 第一个条件是基于(6.11)式：只有 d_n 达到这样的阶，由(6.11)才能得出 $R_n=o(1)$，否则，(6.11)就没有意义了. A_9 的后一条件以及 A_{10}的作用可在后面引理 6.3 的证明中看到.

A_9 对 d_n 给出了两个阶的要求. 哪一个强，哪一个弱，要视 μ_n 而定. 在 $\mu_n\sim n^a$ 时，后一个阶比前一个阶高(对任何 $a>0$).

3）真正本质的条件是 A_7 和 A_8. 它们是对函数 ρ 与 e_i 的公共分布 F 的较强的限制(相对于相合性与渐近正态性的问题而言). 我们先考察几个重要例子，以探究这些条件的含义如何.

例 6.1 对于LADE，此处 $\rho(u)=|u|$，$\Psi(u)=\mathrm{sgn}(u)$，当 $u\neq 0$ 而 $\Psi(0)=\alpha$ 时(对某个 $\alpha\in[0,1]$)，A_7 自然成立. 又 $G(u)=E\Psi(e_1+u)=1-2F(1-u)$. 若 $f=F'$ 在 0 的邻域内存在，则 $G'(u)=2f(-u)$存在，且 $g=2f(0)$. 为使 $g>0$，应要求 $f(0)>0$. 最后，A_8 的后一条件要求

$$|f(u)-f(0)|=O(\sqrt{|u|}). \tag{6.12}$$

总结而言：对 LADE，$A_7\sim A_8$ 要求 F 在其中位数 0 附近有连续非 0 的密度 f，f 满足(6.12).

还应注意：在本例中 $E\Psi'(e_1)=0$，故 g 不是按导出(6.2)时那样取 $h=E\Psi'(e_1)$为值，而是 $2f(0)$. 类似情况在以下还会见到.

例 6.2 Huber 函数

$$\rho(u)=\begin{cases}u^2, & |u|\leqslant c;\\ 2a(u-c)+c^2, & u>c;\\ -2a(u+c)+c^2, & u<-c.\end{cases}$$

c 与 a 为常数，$0<c\leqslant a$ （$a\geqslant c$ 保证了 ρ 为凸函数），分两种情况：

1） $c=a$：这时 $\Psi=\rho'$ 处处存在，有界，且满足 Lip(1)，这时 A_7 当然成立（对这种情况，R_n 的阶可改善，见后面第三段定理6.2）.但为使 $E\Psi(e_1)=0$，对 e_i 的公共分布 F 应有所限制，例如假定 F 关于 0 点对称. 为考察 A_8，设 $|u|$ 充分小，且 F 在 $\pm c$ 点的邻域内有连续密度 f. 易见

$$G(u)=2c[1-F(-u+c)]+2\int_{-u-c}^{-u+c}(x+u)\mathrm{d}F(x)$$
$$-2cF(-u-c),$$

在上述假定下，当 $|u|$ 充分小时，有

$$G'(u)=2F(-u+c)-2F(-u-c).$$

由于 F 在 $\pm c$ 附近有连续密度，由此式知 $|G'(u)-g|=O(|u|^{1/2})$ 的条件满足（实际上为 $O(u)$）. $g>0$ 的条件归结为 $F(c)-F(-c)>0$. 总结而言：对本例的 $c=a$ 的情况，条件 A_7、A_8 归结为：F 在 $\pm c$ 附近有连续密度，$F(c)-F(-c)>0$.

有两点值得注意：虽然在此处 $\rho'=\Psi$ 处处存在，但 $\pm c$ 是 ρ'' 不存在的点，故分布 F 在这些点的邻域内的光滑性有要求；其次，$g=2(F(c)-F(-c))=E\Psi'(e_1)$ 与(6.2)式中的 h 一致：在 ρ' 处处存在时的情况总是如此（当然，F 在 ρ'' 不存在的点的邻域内有要求）；$g\neq E\Psi'(e_1)$ 的情况只出现在 ρ' 非处处存在时.

2） $c<a$：这时 Ψ 仍为有界，A_7 满足. A_8 的情况更复杂一些：当 $|u|$ 充分小时，

$$G'(u)=2(a-c)f(-u+c)+2(a-c)f(-u-c)$$
$$+2F(-u+c)-2F(-u-c).$$

此式当 F 在 $\pm c$ 的邻域内有连续密度 f 时成立. 由此式可见：为了 $g=G'(0)>0$，只须 $f(c)$、$f(-c)$ 和 $F(c)-F(-c)$ 至少有一个不为 0，这比前面要求 $F(c)-F(-c)>0$ 弱一些. 而为 $|G'(u)-g|=O(|u|^{1/2})$，必须 $|f(\pm c+u)-f(\pm c)|=O(|u|^{1/2})$，$u\to 0$（当 $c=a$ 时，无此要求）. 又此处

$$g=2(a-c)(f(c)+f(-c))+2(F(c)-F(-c))$$

$$= 2(a-c)(f(c)+f(-c))+E\Psi'(e_1)$$

一般与 $h=E\Psi'(e_1)$不同，且一般总比 h 大.

这里我们仍然看出：F 在 ρ'或 ρ''不存在的点的领域内，其光滑性有要求：若 ρ'存在但 ρ''不存在，要求少一些；若 ρ'不存在，则要求多一些.

例 6.3 取 $\delta\in(0,1)$，令 $\rho(u)=\delta|u|+(1-\delta)u^2$，这可以看作是 LADE 和 LSE 的一种折衷，希望能以此折衷吸收二者的优点. 为简化讨论，设 e_i 的公共分布 F 关于 0 对称，这时可取 $\Psi(0)=0$. 显见 A_7 满足. 又

$$G(u)=\delta(1-2F(-u))+2(1-\delta)u$$

（此式必须假定 $E|e_i|<\infty$，这时 $Ee_i=0$，因为 F 关于 0 对称），假定 F 在 0 点的邻域内有连续密度 f，则

$$G'(u)=2\delta f(-u)+2(1-\delta),$$

不论 $f(0)$是否大于 0，总有

$$g=2\delta f(0)+2(1-\delta)>0.$$

这与例 6.1 不同. 而 $|G'(u)-g|=O(|u|^{1/2})$的成立需要

$$|f(u)-f(0)|=O(|u|^{1/2}),\ u\to 0.$$

这与例 6.2 中 $a>c$ 的情况相同.

在应用上常见的函数 ρ，多是这种情况：除有限点$a_1<\cdots<a_m$外，ρ''存在. 对这种 ρ，根据以上各例（它显然也适用于一般情况，不限于这几个例子），A_7、A_8 的成立要求以下几条：

1）在 a_i 的邻域内，e_i 的公共分布 F 有连续密度 f；

2）如对其中某个 a_i，$\rho'(a_i)$也不存在，则 f 应进一步满足：

$$|f(a_i+u)-f(a_i)|=O(\sqrt{|u|}),\ u\to 0; \qquad (6.13)$$

3）若 ρ'在 R^1 处处存在（包括 $a_1,\cdots,a_m$），则要求

$$h\equiv E\Psi'(e_1)>0. \qquad (6.14)$$

这时 $G'(0)=g=h$. 若 ρ'在 $a_1,\cdots,a_m$ 中的某些点——例如 $a_1,\cdots,a_j$ 处不存在，于是或者有(6.14)，或者至少对某个 k $(1\leqslant k\leqslant j)$有 $f(a_k)>0$. 在后一情况有 $g\neq h$.

仅仅以上几条尚不足以保证 A_7、A_8 成立，还涉及以下几点：当 $|u|\to\infty$ 时，$\rho(u)$ 增长不能太快，$\rho(u)$ 在 R^1 各部分上的变化不能过于不均匀. 例如，当 $a>2$ 时，$\rho(u)=|u|^a$ 不满足 A_7. 一般地，如果当 $|u|\to\infty$ 时 $|\rho''(u)|\to\infty$，A_7 不能满足. 这限制了 ρ 大体上应是 $O(u^2)$ 级的. 我们不易把一切满足 $A_7\sim A_9$ 的 ρ 简单地刻划出来，但下面的充分条件是简单有用的：除满足前述三条外，若在每一区间 (a_i,a_{i+1}) $(i=0,1,\cdots,m;\ a_0=-\infty,\ a_{m+1}=\infty)$ 内，ρ'' 有界且满足 Lip(1/2)，则 A_7、A_8 成立. 当然，不止限于这种情况.

这些条件似乎对函数 ρ 的光滑性加上了颇高的要求，但从实用观点看，其要害之处在于：它容许 ρ' 和 ρ'' 可以在某些点不存在. 这一点至关重要，它使我们不致把一些应用上重要的 ρ(如 $\rho(u)=|u|$)排除在外.

二、定理 6.1 的证明

证明的方法适用于后面要讨论的其他情况，细节上也大同小异. 因此，我们要仔细写出这一证明来，以方便读者将其推于其他情况. 由于证明甚繁，我们将所需的预备事实及证明中的关键部分列为一些引理如下：

引理 6.1(Berry-Esseen 不等式)　设随机变量 $\xi_1,\cdots,\xi_n$ 相互独立，各有期望 0，$0<B_n^2=\sum_{i=1}^n E\xi_i^2<\infty$，$C_n=\sum_{i=1}^n E|\xi_i|^3$. 用 F_n 和 Φ 分别记 $B_n^{-1}\sum_{i=1}^n\xi_i$ 和 $N(0,1)$ 的分布函数，则存在绝对常数 A，使

$$H_n\equiv\sup_x|F_n(x)-\Phi(x)|\leqslant AC_n/B_n^3\quad(A\text{ 可取为 }0.8).$$

证明见文献[49]中 p. 111.　∎

引理 6.2　在引理 6.1 的记号下，若 $B_n\to\infty$，$B_{n+1}/B_n\to1$，且存在 $a>1$，使 $H_n=O((\log B_n)^{-a})$，则 $\{\xi_i\}$ 满足重对数律：

$$\lim{\sup\atop\inf}\left(\sum_{i=1}^n\xi_i/2B_n^2\log\log B_n^2\right)^{1/2}=\begin{cases}1;\\-1,\end{cases}\quad\text{a.s.}.$$

证明见文献[49]中 p. 305.　∎

引理 6.3　在定理 6.1 的条件下有：

$$\bar{\beta}_n \equiv g^{-1}\sum_{i=1}^{n}\Psi(e_i)x_{ni} = O((\log\log\mu_n)^{1/2}). \quad \text{a.s..} \qquad (6.15)$$

为了证明，以 $x_{i(j)}$ 记 x_i 的第 j 个分量；令 $Q_n=\sum_{i=1}^{n}\Psi(e_i)x_{i(j)}$，$Q_n$ 是 $\sum_{i=1}^{n}\Psi(e_i)x_i$ 的第 j 个分量. 按假定 A_3、A_5 及 A_6，有

$$EQ_n = 0, \ 0 < B_n^2 = \mathrm{Var}(Q_n) = \sigma^2\sum_{i=1}^{n}x_{i(j)}^2 < \infty, \qquad (6.16)$$

以及

$$C_n \equiv \sum_{i=1}^{n}E|\Psi(e_i)x_{i(j)}|^3 = O\Big(\sum_{i=1}^{n}|x_{i(j)}|^3\Big) = O\Big(\max_{1\le i\le n}\|x_i\|\sum_{i=1}^{n}x_{i(j)}^2\Big).$$

注意到

$$\sum_{i=1}^{n}x_{i(j)}^2 \sim \mu_n \quad 及 \quad \max_{1\le i\le n}\|x_i\| = \max_{1\le i\le n}\|S_n^{1/2}x_{ni}\| = O(\mu_n^{1/2}d_n^{1/2}),$$

有
$$C_n = O(d_n^{1/2}\mu_n^{3/2}).$$

又由(6.16)，有 $B_n^2\sim\mu_n$. 于是由引理 6.1 得 $H_n=O(d_n^{1/2})$. 因为 $\lambda_n\sim\mu_n$，有 $d_n\sim\mu_n^{-1}\max_{1\le i\le n}\|x_i\|^2$. 由此，注意到 $d_n\to0$ 及 $\mu_n\sim B_n^2\sim\sum_{i=1}^{n}\|x_i\|^2$，得到 $B_{n+1}/B_n\to1$. 由 $d_n\to0$ 知 $\mu_n\to\infty$，所以 $B_n\to\infty$. 依 $d_n=O((\log\mu_n)^{-\delta})=O((\log B_n)^{-\delta})$ 及上面证明的 $H_n=O(d_n^{1/2})$，有

$$H_n = O((\log B_n)^{-\delta/2}),$$

而 $\delta/2>1$. 这样，引理 6.2 的条件全部满足. 有

$$\lim\begin{matrix}\sup\\ \inf\end{matrix}(Q_n/(2B_n^2\log\log B_n^2)^{1/2}) = \begin{cases}1;\\ -1,\end{cases} \quad \text{a.s..} \qquad (6.17)$$

因为 $\lambda_n\sim\mu_n$，知 $S_n^{-1/2}$ 的各元都是 $O(\mu_n^{-1/2})$，由此及(6.17)以及 $\bar{\beta}_n=g^{-1}S_n^{-1/2}\sum_{i=1}^{n}\Psi(e_i)x_i$，即得(6.15). 引理证毕. ▌

引理 6.4 在定理 6.1 的条件下有：

$$d_n\log\log\mu_n = o(1), \qquad (6.18)$$

$$\log\log\mu_n = O(\log n). \qquad (6.19)$$

为了证明，注意，因为 $\sum_{i=1}^{n}\|x_{ni}\|^2=p$，有 $d_n\geqslant p/n$. 将此与 A_9 的后一式结合，可知存在 $a>0$，使 $(\log\mu_n)^8\leqslant an$，由此即得(6.19). 另一方面，按 A_{10}的前一式，有 $d_n=o(1/\log n)$. 将此与(6.19)结合，得(6.18). 引理证毕. ▌

以下为表述简单计，我们用 c 记一个正常数，其值每次出现时都可以不同，换言之，c 只是常数的符号而不代表数值.

任取 $t_n\to\infty$，但 $t_n d_n^{1/4}(\log n)^{1/2}(\log\log\mu_n)^{1/4}\to 0$，$t_n$ 的存在由条件 A_9 保证. 定义 R^p 中的立方体为：

$$D=D^{(n)}=\{\beta: |\beta|\leqslant(t_n\log\log\mu_n)^{1/2}\},$$

其中 $|\beta|$ 表示 β 诸分量绝对值最大者. 因为 $t_n\leqslant n$，则当 n 充分大时，D 的边长不超过 n. 将 D 剖分为 N 个两两无公共点的相等的立方体 $D_1,\cdots,D_N$，并取 N 使 D_1 的边长 $\nu_n\sim n^{-1}$，则有 $N=O(n^{2p})$. 令

$$\varepsilon_n=t_n d_n^{1/4}(\log n)^{1/2}(\log\log\mu_n)^{1/4}, \tag{6.20}$$

$$\gamma_i(\beta)=\gamma_{ni}(\beta)=\int_0^{-x_{ni}'\beta}(\Psi(e_i+u)-\Psi(e_i)-G(u))\mathrm{d}u \quad (1\leqslant i\leqslant n).$$

引理 6.5　在定理 6.1 的条件下，对任意 l $(1\leqslant l\leqslant N$，l 可以与 N 有关)，当 n 充分大(与上述范围内的 l 无关)时，有

$$P\left(\sup_{\alpha,\beta\in D_l}\left|\sum_{i=1}^{n}(r_i(\beta)-r_i(\alpha))\right|\geqslant\varepsilon_n^2/3\right)\leqslant n^{-c_n}. \tag{6.21}$$

这里 $\{c_n\}$ 为与上述范围内的 l 无关的正数列，且当 $n\to\infty$ 时，$c_n\to\infty$.

为了证明，用 $\overline{D}_l$ 记 D_l 的闭包. 对每个 i $(1\leqslant i\leqslant n)$，可以找到 γ_{i1}和 γ_{i2}属于 $\overline{D}_l$，使

$$-x_{ni}'\gamma_{i1}\leqslant -x_{ni}'\beta\leqslant -x_{ni}'\gamma_{i2},\ \beta\in\overline{D}_l.$$

考虑三种情况：

1° $(x_{ni}'\gamma_{i1})(x_{ni}'\gamma_{i2})<0$：这时必存在 $\gamma_{i0}\in\overline{D}_l$，使 $x_{ni}'\gamma_{i0}=0$. 由 Ψ 非降，易知

$$\left|\int_{-x'_{ni}\alpha}^{-x'_{ni}\beta}(\Psi(e_i+u)-\Psi(e_i))\mathrm{d}u\right|$$
$$\leqslant\sum_{j=1}^{2}\int_{x'_{ni}\gamma_{i0}}^{\cdot x'_{ni}\gamma_{ij}}(\Psi(e_i+u)-\Psi(e_i))\mathrm{d}u. \quad (6.22)$$

2° $-x'_{ni}\gamma_{i2}\leqslant 0$：取 $\gamma_{i0}=\gamma_{i2}$，仍得(6.22).

3° $-x'_{ni}\gamma_{i\cdot}\geqslant 0$：取 $\gamma_{i0}=\gamma_{i1}$，仍得(6.22).

综合 1°～3°，可知总存在 $\gamma_{i0}\in\overline{D}_i$，满足

$$(-x'_{ni}\gamma_{i0})(-x'_{ni}\gamma_{ij})\geqslant 0 \quad (j=1,2), \quad (6.23)$$

使(6.22)成立.

由于 $\|x_{ni}\|\leqslant\sqrt{d_n}$，根据 $t_nd_n\log\log\mu_n\to 0$(由 t_n 的取法及 A_9 推出)，有

$$\sup\{|x'_{ni}\beta|:1\leqslant i\leqslant n,\beta\in D\}\to 0,\ n\to\infty. \quad (6.24)$$

故当 n 充分大时，$|x'_{ni}\gamma_{ij}|$ 一致地充分小. 根据假定 A_8 及 $G(0)=0$，知 $|G(u)|\leqslant c|u|$ 当 $|u|$ 充分小. 故

$$\sum_{i=1}^{n}\left|\int_{-x'_{ni}\gamma_{i0}}^{-x'_{ni}\gamma_{ij}}|G(u)|\mathrm{d}u\right|\leqslant c\sum_{i=1}^{n}\left|\int_{-x'_{ni}\gamma_{i0}}^{-x'_{ni}\gamma_{ij}}|u|\mathrm{d}u\right|. \quad (6.25)$$

考察上式右端的积分，由(6.23)，可知

$$2\left|\int_{x'_{ni}\gamma_{i0}}^{\cdot x'_{ni}\gamma_{ij}}|u|\mathrm{d}u\right|=|(x'_{ni}\gamma_{ij})^2-(x'_{ni}\gamma_{i0})^2|$$
$$\leqslant\|x_{ni}\|^2\|\gamma_{ij}-\gamma_{i0}\|\ \|\gamma_{ij}+\gamma_{i0}\|$$
$$\leqslant c\|x_{ni}\|^2n^{-1}(t_n\log\log\mu_n)^{1/2}. \quad (6.26)$$

最后的不等式是因为 γ_{ij} 与 γ_{i0} 同属于 $\overline{D}_i\subset D$，以及 D_i 的边长 $\sim n^{-1}$. 注意，$\sum_{i=1}^{n}\|x_{ni}\|^2=p$，$\varepsilon_n$ 的定义，$d_n\geqslant 1/n$ 以及 $\log\log\mu_n=O(\log n)$ 和(6.26)式，即得 $\sum_{j=1}^{2}\sum_{i=1}^{n}\left|\int_{-x'_{ni}\gamma_{i0}}^{-x'_{ni}\gamma_{ij}}|G(u)|\mathrm{d}u\right|=o(\varepsilon_n^2)$.

由于 γ_{i1} 和 γ_{i2} 的选法，易见在上式中若把 γ_{i0} 和 γ_{ij} 分别换成 D_i 内任意的 α 和 β，则上式仍成立，且对 D_i 内的 α 和 β 一致成立. 利用这个事实，结合(6.22)，即知当 n 充分大时，有

$$P\left(\sup_{\alpha,\beta\in D_l}\left|\sum_{i=1}^{n}(r_i(\beta)-r_i(\alpha))\right|\geqslant\varepsilon_n^2\right)$$

$$\leqslant\sum_{j=1}^{2}P\left(\sum_{i=1}^{n}(r_i(\gamma_{lj})-r_i(\gamma_{l0}))\geqslant\varepsilon_n^2/3\right).\qquad(6.27)$$

下面的工作是估计上式右边的值. 为此,利用 Bennett 不等式(引理 3.3),暂记 $\xi_i=r_i(\gamma_{lj})-r_i(\gamma_{l0})$,按 $G(u)$的定义及 $E\Psi(e_1)=0$,知 $E\xi_i=0$. 为估计 $\mathrm{Var}(\xi_i)=E\xi_i^2$,利用 Schwartz 不等式,得

$$\xi_i^2\leqslant\left|\int_{x_{ni}'\gamma_{l0}}^{x_{ni}'\gamma_{lj}}\mathrm{d}u\right|\cdot\left|\int_{-x_{ni}'\gamma_{l0}}^{-x_{ni}'\gamma_{lj}}(\Psi(e_i+u)-\Psi(e_i)-G(u))^2\mathrm{d}u\right|,$$

由此得

$$E\xi_i^2\leqslant\|x_{ni}\|\cdot\|\gamma_{lj}-\gamma_{l0}\|$$

$$\times\left|\int_{-x_{ni}'\gamma_{l0}}^{-x_{ni}'\gamma_{lj}}E(\Psi(e_i+u)-\Psi(e_i))^2\mathrm{d}u\right|.$$

往下证明:对充分小的 $l_1>0$,存在只依赖于 l_1 的常数 l_2,使当$|a|\leqslant l_1$,$|b|\leqslant l_1$时,有

$$\left|\int_a^b E(\Psi(e_i+u)-\Psi(e_i))^2\mathrm{d}u\right|\leqslant l_2\left|\int_a^b|u|\mathrm{d}u\right|.$$

事实上,若a、b 同号,则由 Ψ 的非降性,可知$\Psi(e_i+u)-\Psi(e_i)$在u属于上述积分区间内也同号. 又根据条件 A_7 及 Ψ 的非降性,有

$$l_2'\equiv\sup_{|u|\leqslant l_1}\sup_{|t|<\infty}|\Psi(t+u)-\Psi(t)|<\infty.$$

于是得到

$$\left|\int_a^b E(\Psi(e_i+u)-\Psi(e_i))^2\mathrm{d}u\right|$$

$$\leqslant l_2'\left|\int_a^b E(\Psi(e_i+u)-\Psi(e_i))\mathrm{d}u\right|.$$

再用条件 A_8,因为 l_1 充分小,存在 $l_2''>0$,使

$$|E(\Psi(e_i+u)-\Psi(e_i))|\leqslant l_2''|u|\quad(\text{当 }|u|\leqslant l_1\text{ 时}).$$

于是得到

$$\left|\int_a^b E(\Psi(e_i+u)-\Psi(e_i))^2\mathrm{d}u\right|\leqslant l_2'l_2''\left|\int_a^b|u|\mathrm{d}u\right|.$$

取 $l_2=l_2'l_2''$，即得所欲证者. 如果 a 与 b 异号，为确定计，不妨设 $a<0<b$，将$\int_a^b$表成两个积分和$\int_a^0+\int_0^b$，并对每个积分应用上述已证得的结果，也得所欲证者.

再回到前面的证明. 由于$|x_{ni}'\gamma_{i0}|$和$|x_{ni}'\gamma_{ij}|$当 n 充分大时一致地充分小. 用上面证明的事实，即得

$$\begin{aligned}E\xi_i^2&\leqslant c\,\|x_{ni}\|\cdot\|\gamma_{ij}-\gamma_{i0}\|\cdot\left|\int_{-x_{ni}'\gamma_{i0}}^{-x_{ni}'\gamma_{ij}}|u|\mathrm{d}u\right|\\&\leqslant c\sqrt{d_n}\,n^{-2}(t_n\log\log\mu_n)^{1/2}\,\|x_{ni}\|^2.\end{aligned}$$

最后一个不等式是根据(6.26). 因此

$$\sum_{i=1}^n E\xi_i^2\leqslant c\sqrt{d_n}\,n^{-2}(t_n\log\log\mu_n)^{1/2}. \tag{6.28}$$

注意到 $r_i(\beta)$ 的定义，$|x_{ni}'\gamma_{ij}|$ 充分小，$G(u)$ 在 $|u|$ 充分小时有界，以及条件 A_7，有

$$|\xi_i|\leqslant c|x_{ni}'(\gamma_{ij}-\gamma_{i0})|\leqslant c\sqrt{d_n}\,n^{-1}. \tag{6.29}$$

由(6.28)和(6.29)，利用 Bennett 不等式，得

$$\begin{aligned}&P\left(\sum_{i=1}^n(r_i(\gamma_{ij})-r_i(\gamma_{i0}))\geqslant\varepsilon_n^2/3\right)\\&\leqslant 2\exp\left(-\frac{n(\varepsilon_n^2/3n)^2}{2\left(c\sqrt{d_n}\,n^{-3}(t_n\log\log\mu_n)^{1/2}+c\sqrt{d_n}\,n^{-1}\varepsilon_n^2/3n\right)}\right)\\&=2\exp(-A_n).\end{aligned}$$

按 ε_n 的定义，容易验证 $A_n/\log n\to\infty$. 因此，存在 $c_n\to\infty$，使

$$2\exp(-A_n)\leqslant\exp(-c_n\log n).$$

据此及(6.27)，即得(6.21). 且从以上的论证可看出 $\{c_n\}$ 与 l 的取法无关. 引理 6.5 证毕. ▌

定义

$$\bar{r}_i(\beta)=\rho(e_i-x_{ni}'\beta)-\rho(e_i)+x_{ni}'\beta\psi(e_i)-2^{-1}g(0)(x_{ni}'\beta)^2.$$

引理 6.6　在定理 6.1 的条件下有

$$\widetilde{K}_n \equiv \sup\left\{\left|\sum_{i=1}^{n}(\widetilde{r}_i(\beta)-\widetilde{r}_i(\alpha))\right| ; \alpha \in D, \beta \in D, \|\alpha-\beta\| \leqslant \varepsilon_n\right\}$$

$$= o(\varepsilon_n^2), \text{ a.s.}. \tag{6.30}$$

为了证明，任取 $\alpha,\beta \in D$，$|\alpha-\beta| \leqslant 2\varepsilon_n$. 记 $\xi_i = r_i(\beta) - r_i(\alpha)$，有 $E\xi_i = 0$. 按上一引理中的论证，有

$$\begin{aligned} |\xi_i| &\leqslant c\sqrt{d_1}\,\|\alpha-\beta\| \\ &\leqslant c\sqrt{d_n}\,(\|\alpha\| + \|\beta\|) \\ &\leqslant c\sqrt{d_n}\,(t_n \log\log \mu_n)^{1/2}, \end{aligned} \tag{6.31}$$

$$\mathrm{Var}(\xi_i) \leqslant c\sqrt{d_n}\,\varepsilon_n \left|\int_{-x_{ni}'\alpha}^{-x_{ni}'\beta} |u|\,\mathrm{d}u\right|. \tag{6.32}$$

当 $x_{ni}'\alpha$ 和 $x_{ni}'\beta$ 同号时，这个积分值的估计在上一引理的论证中已作过，见(6.26). 但(6.26)最后一式中的 n^{-1} 应改为 ε_n. 若 $x_{ni}'\alpha$ 和 $x_{ni}'\beta$ 异号，则

$$\begin{aligned} 2\left|\int_{-x_{ni}'\alpha}^{-x_{ni}'\beta} |u|\,\mathrm{d}u\right| &= (x_{ni}'\alpha)^2 + (x_{ni}'\beta)^2 \\ &\leqslant 2(x_{ni}'\alpha - x_{ni}'\beta)^2 \\ &\leqslant 2\|x_{ni}\|^2 \cdot \|\alpha-\beta\|^2 \\ &\leqslant 2\|x_{ni}\|^2 \cdot \|\alpha-\beta\|(\|\alpha\| + \|\beta\|) \\ &\leqslant c\|x_{ni}\|^2 \varepsilon_n (t_n \log\log \mu_n)^{1/2}. \end{aligned}$$

结合以上这两种情况，由(6.32)得

$$\sum_{i=1}^{n} \mathrm{Var}(\xi_i) \leqslant c\sqrt{d_n}\,\varepsilon_n^2 (t_n \log\log \mu_n)^{1/2}. \tag{6.33}$$

由(6.31)和(6.33)，应用 Bennett 不等式，得知对任给的 $\varepsilon > 0$，有

$$\begin{aligned} &P\left(\left|\sum_{i=1}^{n}(r_i(\beta) - r_i(\alpha))\right| \geqslant \varepsilon\varepsilon_n^2/3\right) \\ &\qquad \leqslant 2\exp(-c\varepsilon_n^2/\sqrt{d_n}\,(t_n \log\log \mu_n)^{1/2}) \end{aligned}$$

$$\leqslant 2\exp(-ct_n^{3/2}\log n)\leqslant n^{-c_n'} . \tag{6.34}$$

这里 $\{c_n'\}$ 为一串趋于∞的常数，只与 ε 有关，而与 α、β 无关（只要 α、β 都属于 D 且 $|\alpha-\beta|\leqslant 2\varepsilon_n$）.

现有

$$\begin{aligned}J_n\equiv & P\Big(\sup\Big\{\Big|\sum_{i=1}^{n}(r_i(\beta)-r_i(\alpha))\Big|:\alpha\in D,\beta\in D,\|\alpha-\beta\|\\ & \leqslant\varepsilon_n\Big\}\geqslant\varepsilon\varepsilon_n^2\Big)\\ \leqslant & \sum_{j=1}^{N}P\Big(\sup_{\beta\in D_j}\Big|\sum_{i=1}^{n}(r_i(\beta_j)-r_i(\beta))\Big|\geqslant\varepsilon\varepsilon_n^2/3\Big)\\ & +\sum_{k=1}^{N}P\Big(\sup_{\beta\in D_k}\Big|\sum_{i=1}^{n}(r_i(\beta_k)-r_i(\beta))\Big|\geqslant\varepsilon\varepsilon_n^2/3\Big)\\ & +\sum{}^{*}P\Big(\Big|\sum_{i=1}^{n}(r_i(\beta_j)-r_i(\beta_k))\Big|\geqslant\varepsilon\varepsilon_n^2/3\Big)\\ \equiv & 2J_{n1}+J_{n2}.\end{aligned}\tag{6.35}$$

此处 β_j 为立方体 D_j 的中心（$1\leqslant j\leqslant N$），而 $\sum^*$ 表示求和的范围为：

$$1\leqslant j\leqslant N,\quad 1\leqslant k\leqslant N,\quad |\beta_j-\beta_k|\leqslant 2\varepsilon_n.$$

这里我们利用了这样的事实：D_j 的边长 $\sim n^{-1}=o(\varepsilon_n)$. 注意到 $N=O(n^{2p})$，求和号 $\sum^*$ 下的项数不超过 $N^2=O(n^{4p})$. 利用(6.21)和(6.34)，得

$$J_n\leqslant 2n^{2p}n^{-c_n'}+n^{4p}n^{-c_n'}\ ,\ n \text{ 充分大}.$$

由于此，有

$$\sum_{n=1}^{\infty}J_n<\infty.$$

根据 Borel-Cantelli 引理及 ε 的任意性，得

$$\begin{aligned}K_n\equiv & \sup\Big\{\Big|\sum_{i=1}^{n}(r_i(\beta)-r_i(\alpha))\Big|:\alpha\in D,\beta\in D,\|\alpha-\beta\|\leqslant\varepsilon_n\Big\}\\ = & o(\varepsilon_n^2),\ \text{a.s.}.\end{aligned}\tag{6.36}$$

因为　　$\tilde{r}_t(\beta) = r_t(\beta) + \int_0^{-x_{ni}'\beta}(G(u) - gu)\mathrm{d}u,$

有　$\tilde{K}_n \leqslant K_n + \sum_{i=1}^{n}\left|\int_{-x_{ni}'\alpha}^{x_{ni}'\beta}(G(u) - gu)\mathrm{d}u\right| \equiv K_n + W_n.$

$$(6.37)$$

往下证 $W_n = o(\varepsilon_n^2)$：事实上，由条件 A_8，并注意到(6.24)，有

$$G(u) - gu = \int_0^u (G'(v) - g)\mathrm{d}v = O(|u|^{3/2}),$$

故

$$\left|\int_{-x_{ni}'\alpha}^{-x_{ni}'\beta}|G(u) - gu|\mathrm{d}u\right| \leqslant c\left|\int_{-x_{ni}'\alpha}^{-x_{ni}'\beta}|u|^{3/2}\mathrm{d}u\right| \equiv cV. \quad (6.38)$$

当 $-x_{ni}'\alpha$ 与 $-x_{ni}'\beta$ 同号时，有

$$V \leqslant c\left|\,|x_{ni}'\beta|^{5/2} - |x_{ni}'\alpha|^{5/2}\right|;$$

当 $-x_{ni}'\alpha$ 与 $-x_{ni}'\beta$ 异号时，有

$$V \leqslant c\left|x_{ni}'(\beta - \alpha)\right|^{5/2}.$$

不论在何种情况，易见都有

$$V \leqslant c\|x_{ni}\|^2\cdot\|\beta-\alpha\|(\|\beta\|+\|\alpha\|)\|x_{ni}\|^{1/2}(\|\alpha\|^{1/2}+\|\beta\|^{1/2})$$
$$\leqslant c\varepsilon_n\|x_{ni}\|^2(t_n\log\log\mu_n)^{1/2}d_n^{1/4}(t_n\log\log\mu_n)^{1/4}.$$

由此及(6.38)，知为证 $W_n = o(\varepsilon_n^2)$，只须证

$$d_n^{1/4}(t_n\log\log\mu_n)^{3/4} = o(\varepsilon_n).$$

由 ε_n 的定义，$t_n \to \infty$，并注意到(6.19)，知上式必然成立．从而证明了 $W_n = o(\varepsilon_n^2)$．由此及(6.36)、(6.37)，即得(6.30)．引理证毕．▌

定理 6.1 的证明：令

$$x_{ni} = S_n^{-1/2}x_i \quad (1 \leqslant i \leqslant n),$$
$$\tilde{\beta}_0 = S_n^{1/2}\beta_0,$$

而将模型(1.1)改写为

$$Y_i = x_{ni}'\tilde{\beta}_0 + e_i \quad (1 \leqslant i \leqslant n),$$

以 $\tilde{\beta}_n$ 记 $\tilde{\beta}_0$ 的(使用函数 ρ 产生的)M 估计，则有

$$\tilde{\beta}_n = S_n^{1/2}\hat{\beta}_n.$$

不失普遍性,设 $\beta_0 = 0$,有 $\bar{\beta}_0 = 0$, $Y_i = e_i$.

令

$$D^* = \{\beta : \beta \in R^p, |\beta| \leqslant (t_n \log\log \mu_n)^{1/2} - 1\},$$

定义事件 $A_n = \{\bar{\beta}_n \in D^*\}$, $\bar{\beta}_n$ 由(6.18)定义. 因为 $\varepsilon_n \to 0$,故当 n 充分大时,由 $\alpha \in D^*$, $\|\alpha - \beta\| = \varepsilon_n$,可知 $\beta \in D$. 以 $I_n(\cdot)$ 记 A_n 的指示函数.

注意到 $\tilde{r}_i(\beta)$ 的定义,由简单计算证明

$$\sum_{i=1}^n (\tilde{r}_i(\beta) - \tilde{r}_i(\bar{\beta}_n))$$

$$= \sum_{i=1}^n \{\rho(e_i - x'_{ni}\beta) - \rho(e_i - x'_{ni}\bar{\beta}_n)\} - 2^{-1}g \|\beta - \bar{\beta}_n\|^2.$$

根据引理 6.6,由上式得

$$\sup\left\{I_n(\bar{\beta}_n)\left|\sum_{i=1}^n \{\rho(e_i - x'_{ni}\beta) - \rho(e_i - x'_{ni}\bar{\beta}_n)\} - 2^{-1}g\|\beta - \bar{\beta}_n\|^2\right| ; \beta \in R^p, \|\beta - \bar{\beta}_n\| = \varepsilon_n\right\}$$

$$= o(\varepsilon_n^2), \text{ a.s.}. \tag{6.39}$$

而按引理 6.4,有

$$I_n(\bar{\beta}_n) \to 1, \text{ a.s.},$$

因此按(6.39)可知以概率 1 成立:当 n 充分大时,

$$\inf\left\{\sum_{i=1}^n \rho(e_i - x'_{ni}\beta) ; \|\beta - \bar{\beta}_n\| = \varepsilon_n\right\} > \sum_{i=1}^n \rho(e_i - x'_{ni}\bar{\beta}_n).$$

由于 ρ 的凸性,按定理 1.3 的 3°,由上式即推出 $\|\hat{\beta}_n - \bar{\beta}_n\| < \varepsilon_n$. 这就证明了

$$\hat{\beta}_n - \bar{\beta}_n = O(\varepsilon_n), \text{ a.s.}. \tag{6.40}$$

这对任何满足前面所规定的条件的序列 $\{t_n\}$ 都能成立. 因而有

$$\hat{\beta}_n - \bar{\beta}_n = O(d_n^{1/4}(\log n)^{1/2}(\log\log \mu_n)^{1/4}), \text{ a.s.}.$$

不难看出,这正是(6.2)加(6.11). 定理 6.1 证毕. ▌

例 6.4 将定理 6.1 用于 Bahadur 所讨论的位置参数模型

$$Y_i = \beta_0 + e_i, \tag{6.41}$$

取 $\rho(u)=|u|$，β_0 的 M 估计即 LADE 为 $\hat{\beta}_n=\text{med}(Y_1,\cdots,Y_n)$. 假定 e_i 为 iid.，有中位数 0，e_i 在 0 点邻域内有连续密度 f，$f(0)>0$ 且

$$|f(u)-f(0)|=O(\sqrt{|u|})\quad(\text{当 } n\to 0),$$

则据定理 6.1，我们得到 Bahadur 的结果(6.8)+(6.9).

定理 6.1 所给出的剩余项 R_n 的阶(6.11)的精度如何？有无改善的可能？对这个问题，本例提供了一些线索. 因为 Kiefer 于 1967 年在文献[40]中证明了，对于本例，表达式(6.8)中的 R_n 的阶可改善为：

$$R_n=O(n^{-1/4}(\log\log n)^{3/4}),\ \text{a.s.},\tag{6.42}$$

并且这不能再进一步改善了，以此与(6.11)比较，我们可看出以下两点：

1) (6.11)式中 $d_n^{1/4}$ 的指数 1/4，从总体上说，已不能再升高了，甚至要改进为 $O(d_n^{1/4})$ 也不行. 因为不然的话，就与 Kiefer 的结果相矛盾.

2) $(\log n)^{1/2}(\log\log\mu_n)^{1/4}$ 部分可否改善？不能否定其可能性，改善的方向(如果有可能)，明显的候选者是

$$(\log\log n)^{1/2}(\log\log\mu_n)^{1/4}\quad 或者\quad(\log\log\mu_n)^{3/4}.$$

还应当指出，这里说"$d_n^{1/4}$ 的指数 1/4 不能改善"，是指"从总体上"，即指对一切满足定理 6.1 条件的模型同时成立，而不是指对每一个具体的、满足定理 6.1 的条件的模型都不能改善. 我们不加证明地给出一个例子如下：

例 6.5 考虑线性模型(1.1)，其中 $p=1$，而

$$x_i=1\quad(\text{当 } i\neq k!,k\geqslant 1);$$

$$x_i=\sqrt{i}/(\log i)^2$$

$$(\text{当 } i\geqslant 2, i=k! \text{ 对某个 } k\geqslant 2).$$

而 x_1-1. 取 $\rho(u)=|u|$，可以证明，在一定条件下(参看例 6.4)，有

$$R_n=O((\log n)^{-2}(\log\log n)^{1/2}),\ \text{a.s.}$$

而对本例有 $d_n \sim (\log n)^{4}$，当 $n = k!$，$\mu_n \sim n$. 上式可写为

$$R_n = O(d_n^{1/2}(\log\log\mu_n)^{1/2})$$

而(6.11)中的 $d_n^{1/4}$ 改善为 $d_n^{1/2}$. 从本例也看到：之所以会出现这种情况，根子在于极个别的 x_i 可对 d_n 起很大的影响，但不见得一定对 $\hat{\beta}_n$ 有很大的影响.

另外，我们在上面曾提到(6.11)中的 $\log n$ 部分改善为 $\log\log n$ 或 $\log\log\mu_n$ 的可能性，但细察定理 6.1 的证明，不难悟出，这种可能性即使存在，也非目前所用的证法力所能及. 这个证法基于 Bennett 不等式以导致估计(6.34)，因而不能回避在 ϵ_n 的定义中置入 $\sqrt{\log n}$ 这个因子. 考虑到 Bennett 不等式给出的指数界已无可改进，必须另辟蹊径，例如从建立形如 $\sum_{i=1}^{n}\xi_{ni}$ 的量的重对数律入手.

三、$\lambda_n \sim \mu_n$ 且 ψ 满足 Lip(1)时

定理 6.2 假定：

1) 保持 $A_1 \sim A_6$ 和 A_{10}；

2) 删去 A_7，代之以 ψ 满足 Lip(1)(这个条件推出 A_7)；

3) A_8 中的 $O(|\mu|^{1/2})$ 加强为 $O(u)$；

4) 保持 A_9 的后一条件，而将前一条改为

$$d_n = o((\log n \cdot \log\log\mu_n)^{-1}). \qquad (6.43)$$

则(6.2)式中的剩余项 R_n 满足

$$R_n = ((d_n \log n \cdot \log\log\mu_n)^{1/2}), \text{ a.s.}. \qquad (6.44)$$

注意，在此由于 $\rho' = \psi$ 处处存在，故(6.2)式中的 h 即 $E\psi'(e_1)$ 就等于 $g = G'(0)$.

本定理的证明，从方法到细节都与定理 6.1 完全相似，差别只在于两点：一是关于 $|\xi_i|$ 和 $\mathrm{Var}(\xi_i)$ 的估计(ξ_i 的定义与引理 6.6 中的相同)，此处因 ψ 满足 Lip(1)，从而知

$$|\psi(e_i + u) - \psi(e_i) - G(u)| \leqslant c|u|，当 |u| 充分小.$$

故按前面的论证，有

$$|\xi_i| \leqslant \left|\int_{-x'_{ni}\alpha}^{-x'_{ni}\beta} c|u|\mathrm{d}u\right| \leqslant cd_n\varepsilon_n(t_n\log\log\mu_n)^{1/2}$$

以及

$$\mathrm{Var}(\xi_i) \leqslant c\left|\int_{-x'_{ni}\alpha}^{-x'_{ni}\beta} u^2\mathrm{d}u\right| \cdot \left|\int_{-x'_{ni}\alpha}^{x'_{ni}\beta}\mathrm{d}u\right| \leqslant d_n\varepsilon_n^2 \| x_{ni} \|^2 t_n\log\log\mu_n. \tag{6.45}$$

把定理 6.1 证明中 ε_n 修改为

$$\varepsilon_n = t_n^2(d_n\log n \cdot \log\log\mu_n)^{1/2}, \tag{6.46}$$

则前面的一切推理就通行无阻了. 而最后得出的 R_n 的阶就是 (6.46) 定义的 ε_n 去掉 t_n 这个因子——即 (6.43) 式.

另一点的差别在于 (6.37) 式中 $W_n = o(\varepsilon_n^2)$ 的证明上 (ε_n 由 (6.46) 定义). 这要用到 $|G'(u) - g| = O(u)$ (当 $u \to 0$), 得到当 $-x'_{ni}\alpha$ 与 $-x'_{ni}\beta$ 同号时, 有

$$\left|\int_{-x'_{ni}\alpha}^{-x'_{ni}\beta}(G(u) - gu)\mathrm{d}u\right| \leqslant c\left|\int_{-x'_{ni}\alpha}^{-x'_{ni}\beta}|u^2|\mathrm{d}u\right|$$

$$\leqslant c\big|\,|x'_{ni}\beta|^3 - |x'_{ni}\alpha|^3\big| \leqslant c\sqrt{d_n}\,\varepsilon_n \| x_{ni} \|^2 t_n\log\log\mu_n.$$

当 $-x'_{ni}\alpha$ 与 $-x'_{ni}\beta$ 异号时, 这个估计改为

$$c|x'_{ni}(\beta - \alpha)|^3 \leqslant c\sqrt{d_n}\,\varepsilon_n \| x_{ni} \|^2(\| \alpha \| + \| \beta \|)^2$$

$$\leqslant c\sqrt{d_n}\,\varepsilon_n \| x_{ni} \|^2 t_n\log\log\mu_n.$$

二者相同, 不难看出, 由这一估计并利用 (6.19), 即得 $W_n = o(\varepsilon_n^2)$. 正是这一步需要把 A_8 中的 $O(|u|^{1/2})$ 加强为 $O(u)$.

将 (6.44) 和 (6.11) 相比, (6.44) 多出了一个因子 $(d_n\log\log\mu_n)^{1/4}$. 其原因正是在于此处 $\mathrm{Var}(\xi_i)$ 的估计 (6.45) 比以前的估计多出了一个因子 $(d_n\log\log\mu_n)^{1/2}$, 这一点通过 Bennett 不等式的使用, 影响了 ε_n 的选定.

适合本定理的一个例子是例 6.2 的 Huber 函数 $c = a$ 的情况 (e_i 的公共分布满足那里的条件, 且

$$|f(u) - f(0)| = O(|u|^{1/2})$$

要加强为 $|f(u)-f(0)|=O(u)$. 与(6.11)相比,(6.43)总算有了显著的改善,但仍未能摆脱掉 $\sqrt{\log n}$ 这个因子. 关键在于:此处虽然假定了 ρ' 处处存在,但仍容许 ρ'' 在某些点不存在,Huber 函数 $(c=a)$ 就是如此. 在下一节中可看到:若进一步假定 ρ'' 处处存在并满足一定的限制,辅之以其他的条件,则 $\sqrt{\log n}$ 这个因于确实可以排除. 这恐怕已达到所能做到的极限了.

四、一般情况($(\lambda_n\sim\mu_n)$ 不必成立)

定理 6.3　假定:

1) 保持 $A_1\sim A_5$、A_7、A_8,删去 A_6、A_9 和 A_{10};

2) 以下两条件之一成立(此条件包含 A_6):

① 矩母函数 $M(u)=E(\exp u\psi(e_1))$ 在 0 附近存在,

$$d_n=o((\log n)^{-3});$$

② 存在 $\delta\in(0,1]$,使 $E|\psi(e_1)|^{2+2/\delta}<\infty$,

$$d_n=O(n^{-\delta}).$$

则(6.2)式(其中的 h 代之以 $g=G'(0)$)中的剩余项 R_n 满足

$$R_n=O(d_n^{1/4}(\log n)^{3/4}),\ \text{a.s.}. \tag{6.47}$$

注意,由于(6.19)和(6.11)式中的阶总不弱于上式中的阶,这正是补充条件 $\lambda_n\sim\mu_n$ 所带来的收益.

本定理引自文献[57]. 只是此处在条件 2)①的情况下,要求 d_n 为 $o((\log n)^{-3})$,而在文献[57]中只要求 d_n 达到 $O((\log n)^{-1/2})$. 其所以作这一改动,是因为根据(6.47),只有当 d_n 为 $o((\log n)^{-3})$ 时,剩余项 R_n 才能达到 $o(1)$ 的阶.

本定理在证法上与定理 6.1 相似,细节上有一差别,由于免除了条件 A_{10} ,引理 6.3 已不能用了,而需代之以下面的引理:

引理 6.7　在定理 6.3 的条件下有

$$\bar{\beta}_n=O((\log n)^{1/2}),\ \text{a.s.}, \tag{6.48}$$

其中 $\bar{\beta}_n$ 由(6.15)式所定义.

先考虑情况 2)①:因为矩母函数 $M(u)$ 在 0 的邻域内存在,故它必在此范围内有任意阶导数,并且可在期望号下求导,由此,再注意 $E\psi(e_1)=0$,即知 $M'(0)=0$. 由于 $d_n=o((\log n)^{-3})$,知

$d_n \log n \to 0$. 以 x_{nij} 记 x_{ni} 的第 j 个分量，对 $M(\cdot)$ 使用带余项的 Taylor 公式可知当 n 充分大时，有

$$M(x_{nij}\sqrt{\log n}) = 1 + 2^{-1}M''(\theta_{ni})x_{nij}^2\log n$$
$$\leqslant 1 + c\log n x_{nij}^2 \leqslant \exp(c\log n \cdot x_{nij}^2).$$

此处 $c > 0$ 为一常数. 取 $b = c + 2$，应用 Markov 公式，有

$$P\Big(\sum_{i=1}^n x_{nij}\psi(e_i) \geqslant b\sqrt{\log n}\Big) = P\Big(\sum_{i=1}^n \sqrt{\log n}x_{nij}\psi(e_i) \geqslant b\log n\Big)$$
$$\leqslant \exp(-b\log n)\prod_{i=1}^n M(\sqrt{\log n}x_{nij})$$
$$\leqslant \exp(-(b-c)\log n) = n^{-2}.$$

上式中利用了

$$\sum_{i=1}^n x_{nij}^2 = 1.$$

由此，用 Borel-Cantelli 引理，即知以概率 1 当 n 充分大时有

$$\sum_{i=1}^n x_{nij}\psi(e_i) < b\sqrt{\log n}.$$

类似地可证明以概率 1 当 n 充分大时，也有

$$\sum_{i=1}^n x_{nij}\psi(e_i) > -b\sqrt{\log n}.$$

这一结果对 $\sum_{i=1}^n x_{ni}\psi(e_i)$ 的任一分量皆成立. 从而证明了(6.48).

情况 2)②可借助引理 3.1 证明之. 也可以这样证：引用[63]中的一个结果：设常数阵列 $\{a_{ni}: 1 \leqslant i \leqslant n, n \geqslant 1\}$ 满足

$$\sum_{i=1}^n a_{ni}^2 = o(1/\log n),$$

且存在常数 c 及 α $(0 < \alpha \leqslant 1)$，使

$$\max_{1\leqslant i\leqslant n}|a_{ni}| \leqslant cn^{-\alpha}.$$

又设随机变量 $\xi_1, \xi_2, \cdots$ 为 iid.，$E\xi_1 = 0$，$E|\xi_1|^{2/\alpha} < \infty$. 则

$$\sum_{i=1}^n a_{ni}\xi_i \to 0, \text{ a.s.}.$$

证明见文献[63]. 在(6.49)式中取

$$a_{ni} = x_{nij}/(\sqrt{\log n}\cdot c_n) \quad (1\leqslant i\leqslant n),$$

其中 c_n 是任一串趋于∞的常数,取 $\xi_i=\psi(e_i)$,按 2)②的假定,易验证此处的条件全成立,因而有

$$\sum_{i=1}^{n} x_{nij}\psi(e_i)/(\sqrt{\log n}c_n) \to 0, \text{ a. s.}.$$

因此式对任何 $c_n\to\infty$ 皆成立,故必有

$$\sum_{i=1}^{n} x_{nij}\psi(e_i) = O(\sqrt{\log n}).$$

引理证毕. ▌

有了这个引理,定理 6.1 的证明就可以全部移植于此处来证明定理 6.3. 唯一要变更的是:在该证明中凡提到 $\log\log\mu_n$ 之处,都要用 $\log n$ 取代. 例如 ε_n 的定义修改为

$$\varepsilon_n = t_n d_n^{1/4}\sqrt{\log n}(\log n)^{1/4} = t_n d_n^{1/4}(\log n)^{3/4},$$

等等. 最后结果(6.11)中的 $\log\log\mu_n$ 也以 $\log n$ 取代,从而得到(6.47). 定理证毕. ▌

在 ψ 满足 Lip(1)时,也有与定理 6.2 相应的结果,见以下定理:

定理 6.4 保持定理 6.3 的假定条件 2)(但情况①中的 $d_n=o((\log n)^{-3})$ 可减弱为 $d_n=o((\log n)^{-2})$),而将其假定条件 1)修改为:保持 $A_1\sim A_5$,删去 A_6、A_9、A_{10},而将 A_8 中的 $|G'(u)-g|=O(|u|^{1/2})$ 加强为 $|G'(u)-g|=O(u)$. 再将 A_7 代之以 ψ 满足 Lip(1). 则(6.2)式中的剩余项 R_n 满足

$$R_n = O(d_n^{1/2}\log n), \text{ a. s.}. \tag{6.49}$$

证明也是基于引理 6.7,其余与定理 6.1 无异. 又,此处因 ρ' 处处存在,故(6.2)中的 h 与 $g=G'(0)$ 是一致的.

五、弱表示

定理 6.5 在定理 6.1 的条件 $A_1\sim A_5$ 和 A_7、A_8 之下,只要 $d_n\to 0$,就有

$$R_n = O_p(d_n^{1/4}(\log d_n^{-1})^{1/2}).$$

又若将条件 A_8 中的数量级 $O(|u|^{1/2})$ 加强为 $O(u)$,则有

$$R_n = O_p(d_n^{1/2}(\log d_n^{-1})^{1/2}).$$

证　证明的方法与定理 6.1 相似，在细节上有几处不同之点．这些不同点是导致本定理中 R_n 的数量级与以前诸定理有所差异的原因所在：

1）(6.15)和(6.48)由

$$\bar{\beta}_n = O_p(1) \tag{6.50}$$

所代替．此式的成立是因为

$$E\|\bar{\beta}_n\|^2 = g^{-2}p\sigma^2 < \infty,$$

且与 n 无关．这样就免除了条件 A_6、A_9、A_{10}及定理 6.3 中的条件 2)．它们的引入是为了证明(6.15)或(6.48)．

2）由于这一变化，立方体 $D = D^{(n)}$ 的定义相应地修改为

$$D = \{\beta: |\beta| \leqslant t_n^{1/2}\},\ t_n \to \infty;$$

ε_n 的定义修改为

$$\varepsilon_n = t_n d_n^{1/4}(\log d_n^{-1})^{1/2},$$

与 ε_n 原来的定义(6.20)相比，去掉了 $(\log n)^{1/2}(\log\log\mu_n)^{1/4}$，而增加了 $(\log d_n^{-1})^{1/2}$ 这个因子．去掉 $(\log\log\mu_n)^{1/4}$ 是因为(6.50)代替了(6.15)，这是不难理解的．至于 $(\log n)^{1/2}$ 和 $(\log d_n^{-1})^{1/2}$ 的一删一增，原因在于：在弱表示的情况我们只须证明(6.21)式左端及(6.34)式左端收敛于 0，而不需要它是收敛级数的一般项(这一点与下面的第 3 条结合起来看)．

3）将 D 剖分为一些相等的立方体 $D_1,\cdots,D_N$ 时，D_i 的边长取为 $\nu_n \sim d_n^{1/3}$，这样 $N = O(t_n^{p/2}d_n^{-p/3})$．$\varepsilon_n$ 中所以要置入因子 $(\log d_n^{-1})^{1/2}$，就是为了抵消这个因子而作的，由于 $d_n \geqslant p/n$，总有 $\log d_n^{-1} = O(\log n)$．

与强表示的情况相似，这里所得到的弱表示的阶中包含了一个对数因子，按以上的证法，这个因子无法免除，然而，采用更复杂一点的证法(同时对条件加以适当的修补)，的确可消除这个因子，从而大大改善上述结果．这个成果包含在 Bai，Rao 和 Zhao 的工作——文献[15]中，该文讨论的是更一般的多重多元线性回归模

型. 以下只介绍其用于模型(1.1)的情况,因篇幅所限,只引述其结论,而略去其证明.

定理 6.6　沿用前而的记号,假定:

1° 条件 $A_1 \sim A_5$ 和 A_7 成立;

2° 存在 $\lambda \in (0,1]$, $g > 0$, 使当 $u \to 0$ 时,有

$$G(u) = gu + O(|u|^{1+\lambda});$$

3° $E(\psi(e_1 + u) - \psi(e_1))^2 = O(|u|^{2\lambda})$;

4° $d_n \to 0$.

则有

$$R_n = O_p(d_n^{\lambda/4}).$$

更进一步,如果把 1°中的"A_7 成立"加强为:存在常数 $\alpha > 0$, $\Delta > 0$ 和 $M < \infty$, 使

$$|\psi(u + v) - \psi(u)| \leqslant M|v|^\alpha, \text{当 } |v| \leqslant \Delta, |u| < \infty \text{ 时},$$

并把 3°加强为:存在 $\Delta > 0, M < \infty$, 使当

$$|u_1| \leqslant \Delta, |u_2| \leqslant \Delta$$

时,有

$$E|\psi(e_1 + u_1) - \psi(e_1 + u_2)|^2 \leqslant M|u_1 - u_2|^{2\lambda}.$$

从而有

$$R_n = O_p(d_n^{\lambda/2}).$$

本定理的条件与定理 6.5 的条件相比较互有出入:A_8 略强于 2°(当 $\lambda = 1/2$ 和 $\lambda = 1$ 时),而此处增加了条件 3°(文献[15]中有一个条件在此处没有列入. 此条件要求:记 ρ 不可微的点所构成的集为 D, 则 $F(D) = 0$, F 为 e_i 的公共分布. 对此处我们所讨论的一重线性回归,即模型(1·1)而言,此条件易由 2°或 3°推出). 另外,本定理容许 λ 取(0,1]中的任意值,而定理 6.5 只相当于本定理的 $\lambda = 1/2$ 和 $\lambda = 1$ 的情况.

定理 6.5 中 R_n 的阶的主要部分 $d_n^{1/4}$ 或 $d_n^{1/2}$ 一般不能再改善,这在强表示的情况已有例证,但强表示的结果不能自动移植于此处,因为一串随机变量的强阶与弱阶可以不一致. 指数 1/4 或 1/2 不能被一般地改善,需要通过具体例子来证实. 对 1/2 的情况的例

子见 § 6.3 中定理 6.7 证明的尾段(在该处的例子中 ρ 为非凸函数,但沿用该例的想法,不难造出 ρ 为凸函数的例子);对 1/4 的情况的例子,可以举 Bahadur 讨论的用样本中位数估计总体中位数(参看 Kiefer 的工作[40]),下面就讨论这个例子.

回到例 6.4,不失普遍性,设 $\beta_0=0$, e_i 为 iid. 并且有公共分布 $R(-1,1)$. 因为

$$\hat{\beta}_n=\mathrm{med}(Y_1,\cdots,Y_n)=\mathrm{med}(e_1,\cdots,e_n),$$

记

$$\xi_n=S_n^{1/2}\hat{\beta}_n=\sqrt{n}\,\hat{\beta}_n,$$

有

$$\xi_n\xrightarrow{\mathscr{L}}N(0,1/4),\ 当\ n\to\infty.$$

又

$$\tilde{\beta}_n=\sum_{i=1}^{n}\mathrm{sgn}(e_i)/\sqrt{n}.$$

以下为避免一些无谓的枝节,设 n 为奇数,而 $N=(n-1)/2$.

现在要利用次序统计量条件分布的一个结果(参看文献[3]中的 p. 13):设 $X_1\leqslant\cdots\leqslant X_n$ 为从一维分布 F 中抽出的 iid. 样本的次序统计量,设 a 为常数,定义分布函数:

$$F_1(x)=\min(F(x)/F(a),1),$$

$$F_2(x)=\max\left(\frac{F(x)-F(a-0)}{1-F(a-0)},0\right),$$

则在给定 $X_m=a$ 的条件下,$(X_1,\cdots,X_{m-1})$ 的条件分布等于从分布 F_1 中抽出的样本量为 $m-1$ 的 iid. 样本的次序统计量的分布. $(X_{m+1},\cdots,X_n)$ 的条件分布与此类似,只须把上文的 F_1 改为 F_2(且这两部分条件独立. 这一点此处不需用到).

将这一结果用于此处,记

$$\eta_n=\sum_{i=1}^{n}\mathrm{sgn}(e_i)/\sqrt{n}-\xi_n,$$

$$\xi_n=\sqrt{n}\,\hat{\beta}_n.$$

并以 $B(N,p)$ 记一随机变量,其分布为

$$P(B(N,p)=i)=\binom{N}{i}p^i(1-p)^{N-i}\quad(i=0,1,\cdots,N).$$

暂设 $a\geqslant 0$，在给定 $\xi_n=a$ 的条件下，$\sum_{i=1}^{n}\mathrm{sgn}(e_i)$ 的条件分布与 $2B\left(N,\frac{a/\sqrt{n}}{1+a/\sqrt{n}}\right)$ 相同. 故在给定 $\xi_n=a$ 的条件下，$n^{1/4}\eta_n$ 的条件分布与

$$\theta_n\equiv n^{1/4}\left(2B\left(N,\ \frac{a/\sqrt{n}}{1+a/\sqrt{n}}\right)\Big/\sqrt{n}-a\right)$$

相同. 使用中心极限定理，或直接计算 θ_n 的特征函数，易见

$$\theta_n\xrightarrow{\mathscr{L}}N(0,2a).$$

若 $a<0$，只须将 $N(0,2a)$ 中的 a 改为 $-a$ 即可.

由于

$$2\xi_n\xrightarrow{\mathscr{L}}N(0,1),$$

从上面的讨论可知，当 $n\to\infty$ 时，$n^{1/4}\eta_n$ 有极限分布，它就是 $|Z_1|Z_2$ 的分布，其中 Z_1 和 Z_2 独立并且各有分布 $N(0,1)$. 此分布的密度函数 h 容易计算得到

$$h(u)=2\pi^{-1}\int_0^{\infty}v^{-1}\exp(-u^2/2v^2-v^2/2)\mathrm{d}v\quad(-\infty<u<\infty).$$

这证明了在本例中，η_n 达不到 $o_p(n^{-1/4})=o_p(d_n^{1/4})$.

总结以上，可知对定理 6.6 的 $\lambda=1/2$ 和 $\lambda=1$ 这两种情况所得到的阶是确切的，一般地不能再有任何改进. 当然，对一个特定的模型不必如此，这在例 6.5 中已见到了.

§ 6.3　ρ不必为凸的情况

当 ρ 不为凸函数时，定理 1.3 的 3°不成立. 这使得问题的困难程度大大增加. 为在 ρ 为非凸的情况下得到 M 估计的线性表示所需的条件要加强，这包括了对 ρ 的光滑性更高的要求、$\{x_i\}$ 的有

界性及限制参数空间为有界等等. 本节的目的就是在这种加强了的条件下,去探讨 M 估计的线性表示的问题.

一、定理的表述

一切沿用 § 6.2 的记号. 有一点不同之处是,在此处我们假定未知参数 β_0 落在 R^p 的一个有界闭区域 B 内. 与此相应, β_0 的 M 估计 $\hat{\beta}_n$ 也就定义为使

$$\sum_{i=1}^{n}\rho(Y_i - x_i'\hat{\beta}_n) = \inf\left\{\sum_{i=1}^{n}\rho(Y_i - x_i'\beta);\beta \in B\right\} \tag{6.51}$$

成立的一个解. 假定:

1° $e_1, e_2, \cdots$ 为 iid.;

2° ρ 非负, $\psi = \rho'$ 及 ψ' 在 R^1 处处存在, ψ' 满足 Lip(1), 且有界;

3° $E|\rho(e_1)|^k < \infty$ 对任何 $k > 0$, 又

$$\sigma^2 \equiv E\psi^2(e_1) > 0,\ g \equiv E\psi'(e_1) > 0;$$

4° 定义

$$h(t) = E\rho(e_1 + t),\ -\infty < t < \infty.$$

则 0 是 $h(t)$ 的唯一最小值点;

5° $\{x_i\}$ 有界, 且存在 $\delta > 0$, 使

$$d_n = O(n^{-\delta}).$$

定理 6.7　设 B 为 R^p 中的有界闭域, 参数真值 β_0 是 B 的一个内点, β_0 的 M 估计 $\hat{\beta}_n$ 由(6.51)定义, 则在条件 1°～5°成立时, (6.2)中的剩余项 R_n 满足

$$R_n = O(d_n^{1/2}\log n),\ \text{a.s.}, \tag{6.52}$$

$$R_n = O_p(d_n^{1/2}). \tag{6.53}$$

其中 d_n 的定义如 § 6.2 开始处所示. 在此, (6.53)中的数量级一般已无可改进.

又若 $\lambda_n \sim \mu_n$ ("～"的意义也见 § 6.2 开始处), 则(6.52)可改进为

$$R_n = O(d_n^{1/2}\log\log n),\ \text{a.s.}. \tag{6.54}$$

将本定理的结论与 § 6.2 中的结果比较, 可以看出:

(1) 在不假定 $\lambda_n \sim \mu_n$ 时,(6.52)与定理 6.4(其中假定 ψ 满足 Lip(1))的结论一样.

(2) 在假定 $\lambda_n \sim \mu_n$ 时,(6.54)达到的数量级优于在 §6.2 中的结果,包括在假定 ψ 满足 Lip(1)及 $\lambda_n \sim \mu_n$ 时的结果,这一改善是由于对 ρ 假定了更严的光滑性.

(3) 弱表示(6.53)已达到最优,这也是 §6.2 中的结果所不可比拟的. 比较此处的条件与 §6.2 的条件,除此处 ρ 不必为凸函数以外,其余条件都比 §6.2 强. 从应用的角度分析,诸如 $\{x_i\}$ 和 B 有界及 $\rho(e_i)$ 的各阶矩存在等等,不算严重的限制. 要害之处在于此处要求 ρ' 和 ρ'' 处处存在(在 §6.2 中,ρ' 可以在个别点不存在),这排除了一些在应用上有兴趣的 ρ. 因此,如何在不假定 ρ 为凸函数并且 ρ' 不必处处存在时去研究 $\hat{\beta}_n$ 的线性表示,是一项有待解决的重要问题.

二、预备事实

引理 6.8 当 $\{x_i\}$ 有界时,有

$$d_n \sim \lambda_n^{-1}. \tag{6.55}$$

首先,$d_n = O(\lambda_n^{-1})$ 容易通过把 S_n 化为对角形并利用 $\{x_i\}$ 有界而推出. 其次,$\lambda_n^{-1} = O(d_n)$ 由(3.11)得出. ▌

注意 后一部分的证明未用到 $\{x_i\}$ 有界. 至于前一部分,易见没有 $\{x_i\}$ 有界的假定是不行的.

引理 6.9 在定理 6.7 的条件下有:

1° 在条件 4°中定义的 $h(t)$ 在 R^1 上处处有限、连续.

2° $E\psi(e_1) = 0$ 且 $|\psi(e_1)|$ 的任意阶矩有限.

3° 对给定的常数 $L > 0$,存在只与 L 有关的常数 $c_1 > 0$, $c_2 > 0$,使

$$h(t) - h(0) \geqslant c_1 t^2,\ |t| \leqslant L, \tag{6.56}$$

$$\mathrm{Var}(\rho(e_1 + t) - \rho(e_1)) \leqslant c_2 t^2,\ |t| \leqslant L. \tag{6.57}$$

4° 对任意给定的常数 $m > 0$, $L > 0$,有

$$\sup\{E|\rho(e_1 + t) - \rho(e_1)|^m : |t| \leqslant L\} < \infty. \tag{6.58}$$

为了证明,作 Taylor 展开:

$$\rho(e_1 \pm t) = \rho(e_1) \pm t\psi(e_1) + 2^{-1}t^2\psi'(e_1 \pm \theta_{\pm}t),\ |\theta_{\pm}| \leqslant 1. \tag{6.59}$$

根据假定，$M \equiv \sup_x |\psi'(x)| < \infty$. 将(6.59)中的两式相加，注意到函数 ρ 为非负，得

$$\rho(e_1 + t) \leqslant \rho(e_1 + t) + \rho(e_1 - t) \leqslant 2\rho(e_1) + Mt^2. \tag{6.60}$$

由假定 3°，即可知 $\rho(e_1 + t)$ 有任意阶的有限矩. 特别是 $h(t) = E\rho(e_1 + t)$ 有限. 在(6.59)中取 $t = 1$ 并取 $+$ 号，得

$$|\psi(e_1)| \leqslant \rho(e_1) + \rho(e_1 + 1) + M.$$

由此及假定 3°，知 $|\psi(e_1)|$ 的任意阶矩有限. 利用这一点以及由(6.59)得出的

$$|\rho(e_1 + t) - \rho(e_1)| \leqslant |t\psi(e_1)| + Mt^2,$$

即得(6.57).

$h(t)$ 的连续性由(6.60)，利用 ρ 的连续性和控制收敛定理可推出来，由(6.59)有

$$|(h(t) - h(0))/t - E\psi(e_1)| \leqslant M|t|.$$

令 $t \to 0$，得 $h'(0) = E\psi(e_1)$. 因为 0 是 $h(t)$ 的最小值点，故有 $h'(0) = 0$. 这证明了 $E\psi(e_1) = 0$. 利用 $E\psi(e_1) = 0$ 及(6.59)，即得

$$h(t) - h(0) = 2^{-1}t^2E\psi'(e_1 + \theta_+t). \tag{6.61}$$

因为 ψ' 连续且有界，由控制收敛定理，有

$$\lim_{t\to 0} E\psi'(e_1 + \theta_+t) = E\psi'(e_1) = g > 0.$$

由此与(6.61)，得

$$\lim_{t\to 0}(h(t) - h(0))/t^2 = g/2 > 0.$$

再结合当 $t \neq 0$ 时，$h(t) > h(0)$，即得(6.56). 最后，(6.58)根据(6.60)及 $\rho(e_1)$ 的各阶矩有限的假定即可推出来. 引理证毕. ▌

引理 6.10　在本定理条件下，(6.48)成立. 若 $\lambda_n \sim \mu_n$，则(6.15)成立，且其中的 μ_n 可改为 n.

事实上，证明(6.48)或(6.15)所用的推理在此处的假定下全部有效. 又若 $\mu_n \sim \lambda_n$，由引理 6.8，有 $\mu_n \sim d_n^{-1}$. 结合条件 5°，知存在常数 $c > 0$，使当 n 充分大时，有 $\mu_n \geqslant cn^{-\delta}$. 由此知

$$\log\log n = O(\log\log\mu_n).$$

另一方面，由于 $\{x_i\}$ 有界，知存在常数 $c' > 0$，使 $\mu_n \leqslant c'n$. 于是又有

$$\log\log\mu_n = O(\log\log n).$$

这样便得到

$$\log\log\mu_n \sim \log\log n,$$

从而在 $\lambda_n \sim \mu_n$ 之下，(6.15)式中的 $\log\log\mu_n$ 可改为 $\log\log n$. 引理证毕. ▌

引理 6.11 仍如前，记

$$x_{ni} = S_n^{-1/2}x_i \quad (1 \leqslant i \leqslant n),$$

并定义

$$A_n = \sum_{i=1}^{n}(\psi'(e_i) - g)x_{ni}x_{ni}'. \tag{6.62}$$

则在定理 6.7 的条件下，有

$$A_n = O((d_n\log n)^{1/2}), \text{ a.s.}. \tag{6.63}$$

又若 $\lambda_n \sim \mu_n$，则(6.63)可改进为

$$A_n = O((d_n\log\log n)^{1/2}), \text{ a.s.}. \tag{6.64}$$

暂且以 a_{ni} 和 b_{ni} 分别记 x_{ni} 的第 j 个和第 k 个分量($j,k = 1, \cdots, p$). (6.63)的意义是

$$\eta_n \equiv \sum_{i=1}^{n}(a_{ni}b_{ni})\xi_i = O((d_n\log n)^{1/2}), \text{ a.s.}. \tag{6.65}$$

其中 $\xi_i = \psi'(e_i) - g$，有 $E\xi_i = 0$ 且 ξ_i 有界(假定 2°). 任取常数串 $c_n \to \infty$，而考虑 $\eta_n/(c_n(d_n\log n)^{1/2})$. 不难验证，在定理 6.7 的假定下，引理 6.7 证明中所引用的文献[63]中那个结果适用于 $\eta_n/(c_n(d_n\log n)^{1/2})$. 按该结果，可得到

$$\eta_n/(c_n(d_n\log n)^{1/2}) \to 0, \text{ a.s.}.$$

由于这一事实对任何 $c_n \to \infty$ 都成立，故有

$$\eta_n = O((d_n \log n)^{1/2}), \text{ a.s.}.$$

从而证明了(6.63).

(6.64)的证明较复杂一些，仍以 ξ_i 记 $\psi'(e_i) - g$，以 a_i 和 b_i 分别记 x_i 的第 j 个和第 k 个元 $(j,k = 1,\cdots,p)$，记

$$\eta_n = \sum_{i=1}^{n} a_i b_i \xi_i,$$

有 $E\xi_i = 0$. 若 $P(\xi_i = 0) = 1$，则 A_n 以概率 1 为 0，(6.64)当然成立，若不然，就有 $\sigma_0^2 \equiv E\xi_i^2$ 非 0，有限. 记

$$B_n^2 = \sum_{i=1}^{n} \sigma_0^2 a_i^2 b_i^2,$$

分两种情况：

① $B_n^2 \to \infty$：因 $\{x_i\}$ 有界，由此知 $B_{n+1}/B_n \to 1$. 按引理 6.3 的推理，有

$$H_n \equiv \sup |F_n(x) - \Phi(x)| = O(d_n^{1/2}) = O(n^{-\delta/2}).$$

又因 $\{x_i\}$ 有界，有

$$\log B_n = O(\log n),$$

故

$$n^{-\delta/2} = O((\log B_n)^{-c})$$

对任何 $c > 0$ 成立. 这证明了对 η_n 而言引理 6.2 的条件全满足. 得到

$$\eta_n = O((B_n^2 \log\log B_n^2)^{1/2}), \text{ a.s.}. \tag{6.66}$$

现在由 $\lambda_n \sim \mu_n$ 易见（c 为常数）：

$$\begin{aligned} B_n^2 &\leqslant c \max_{1 \leqslant i \leqslant n} \| x_i \|^2 \sum_{i=1}^{n} \| x_i \|^2 \\ &= c \max_{1 \leqslant i \leqslant n} \| S_n^{1/2} x_{ni} \|^2 \operatorname{tr}(S_n) \\ &= O((\lambda_n^{1/2} d_n^{1/2})^2) O(\lambda_n) = O(\lambda_n^2 d_n) = O(n^2). \end{aligned}$$

最后一式子是因为 $\{x_i\}$ 有界，故 $\lambda_n = O(n)$，而 $d_n \leqslant p$，由此及(6.66)，得

$$\eta_n = O(\lambda_n d_n^{1/2} (\log\log n)^{1/2}), \text{ a.s.}.$$

这一结论对任何 (j,k) 成立 $(1\leqslant j,\ k\leqslant p)$，故

$$D_n\equiv\sum_{i=1}^{n}(\psi'(e_i)-g)x_ix_i' = O(\lambda_n d_n^{1/2}(\log\log n)^{1/2}),\ \text{a.s.} \tag{6.67}$$

因为
$$A_n=S_n^{-1/2}D_nS_n^{1/2},$$
而
$$S_n^{-1/2}=O(\lambda_n^{-1/2}),$$
由上式即得(6.64).

② $\{B_n^2\}$ 有界：对这个情况，要引用下面的结果：

设 $\xi_1,\xi_2,\cdots$ 为 iid., $E\xi_1=0$, $E\xi_1^{2r}<\infty$ 对某个自然数 r，则

$$\sup\left\{E\left(\sum_{i=1}^{n}a_i\xi_i\right)^{2r}:\sum_{i=1}^{n}a_i^2=1,\ n\geqslant 1\right\}<\infty. \tag{6.68}$$

证明见文献[6]. 由于 $\{B_n\}$ 及 ξ_1 有界，且 $E\xi_1=0$，由(6.68)有 $E\eta_n^{2r}=O(1)$，故当 $r\geqslant 2/\delta$ 时，

$$E(\sqrt{d_n}\,\eta_n)^{2r}=O(d_n^{\,r})=O(n^{-\delta r})=O(n^{-2}).$$

于是

$$\sqrt{d_n}\,\eta_n=o(1),\ \text{a.s.}.$$

这对任何 (j,k) 都成立 $(1\leqslant j,k\leqslant p)$，故 $D_n=o(d_n^{-1/2})$, a.s., D_n 的定义见(6.67). 最后，由

$$D_n=o(d_n^{-1/2}),\ \text{a.s.},$$
$$A_n=S_n^{-1/2}D_nS_n^{-1/2}$$

以及

$$S_n^{-1/2}=O(\lambda_n^{-1/2})=O(d_n^{1/2})\ \text{（见引理 6.8）},$$

有

$$A_n=o(d_n^{1/2}d_n^{-1/2}d_n^{1/2})=o(d_n^{1/2})=o(d_n^{1/2}(\log\log n)^{1/2}),\ \text{a.s.}.$$

即(6.64). 引理证毕. ▌

引理 6.12 在定理 6.7 的条件下，对任给的 $\alpha>0$，有

$$P(\|S_n^{1/2}(\hat{\beta}_n-\beta_0)\|\geqslant n^{\alpha},\ \text{i.o.})=0. \tag{6.69}$$

不失普遍性，设 $\beta_0=0$，因而 $Y_i=e_i$. 记

$$B_n = B \cap \{\beta: \| S_n^{1/2}\beta \| \geqslant n^{\alpha}\}.$$

把 R^p 剖分为一些边长为 n^{-3} 的超立方体：

$$J(i_1,\cdots,i_p) = \{(u_1,\cdots,u_p): l_i n^{-3} \leqslant u_i < (l_i+1)n^{-3}, 1 \leqslant i \leqslant p\},$$

$$l_i = 0, \pm 1, \pm 2, \cdots \quad (1 \leqslant i \leqslant p).$$

若 $J(i_1,\cdots,i_p) \cap B_n \neq \varnothing$，则从此交集中挑出一个点 $\beta_{i_1,\cdots,i_p} \equiv \beta^*$，所有被挑出的点的集记为 G_n. 由于集 B 有界，G_n 中所含的点数 $\#(G_n) \leqslant n^{3p+1}$ （当 n 充分大时）. 记

$$\xi_i = \rho(e_i - x_i'\beta) - \rho(e_i) - (h(-x_i'\beta) - h(0)) \quad (i=1,2,\cdots),$$

$$\begin{aligned}\Delta_n(\beta) &= \sum_{i=1}^{n}(\rho(e_i - x_i'\beta) - \rho(e_i)) \\ &= \sum_{i=1}^{n}\xi_i + \sum_{i=1}^{n}(h(-x_i'\beta) - h(0)) \\ &\equiv \sum_{i=1}^{n}\xi_i + b_n(\beta),\end{aligned}$$

$$V_n(\beta) = \sum_{i=1}^{n}\mathrm{Var}(\rho(e_i - x_i\beta) - \rho(e_i)).$$

因为 $\{x_i\}$ 和 B 有界，知 $\{x_i'\beta: i \geqslant 1, \beta \in B\}$ 有界. 故按(6.56)、(6.57)及 B_n 的定义，知存在常数 $h_1 > 0$、$h_2 > 0$，使对任何 $\beta \in B_n$，有

$$b_n(\beta) \geqslant h_1 \sum_{i=1}^{n}(x_i'\beta)^2 = h_1 \| S_n^{1/2}\beta \|^2 \geqslant h_1 n^{2\alpha}, \tag{6.70}$$

$$V_n(\beta) \leqslant h_2 \sum_{i=1}^{n}(x_i'\beta)^2 = h_2 \| S_n^{1/2}\beta \|^2. \tag{6.71}$$

取 $m \geqslant 2$，其值将在后面选定. 按(6.58). 有

$$\sup_{n \geqslant 1} E|\xi_n|^m \equiv a_m < \infty.$$

利用引理 3.1，并注意(6.70)和(6.71)，有

$$\begin{aligned}&P(\Delta_n(\beta) \leqslant 2^{-1}h_1 n^{2\alpha}) \leqslant P(\Delta_n(\beta) \leqslant 2^{-1}h_1 \| S_n^{1/2}\beta \|^2) \\ &\quad \leqslant P\left(\sum_{i=1}^{n}\xi_i \leqslant -2^{-1}h_1 \| S_n^{1/2}\beta \|^2\right) \leqslant c_1(2^{-1}h_1 n^{2\alpha})^{-m} n a_m \\ &\qquad + 2\exp(-c_2 h_1^2 \| S_n^{1/2}\beta \|^4 / (4h_2 \| S_n^{1/2}\beta \|^2))\end{aligned}$$

$$\leqslant \tilde{c}_1 n^{1-2am} + 2\exp(-\tilde{c}_2 n^{2a}).$$

此处 c_1、c_2、$\tilde{c}_1$、$\tilde{c}_2$ 都是只与 m 有关的正常数. 由此式,并注意到 $\#(G_n) \leqslant n^{3p+1}$,有

$$P\left(\inf_{\beta\in G_n} \Delta_n(\beta) \leqslant 2^{-1} h_1 n^{2a}\right) \leqslant \tilde{c}_1 n^{-2am+3p+2} + 2n^{3p+1}\exp(-\tilde{c}_2 n^{2a}). \tag{6.72}$$

现设 β 是 B_n 中的任一点,找包含此点的立方体 $J(l_1,\cdots,l_p)$. 我们前面已把 G_n 与 $J(l_1,\cdots,l_p)$ 的公共点记为 β^* ,有

$$\| \beta^* - \beta \| \leqslant \sqrt{p}\, n^{-3}.$$

因为 $\{x_i\}$ 、B 和 ψ' 都有界,把函数

$$\rho(e_i - x_i'\beta) - \rho(e_i - x_i'\beta^*)$$

在点 $e_i - x_i'\beta$ 处作 Taylor 展开,有

$$\begin{aligned}
&\rho(e_i - x_i'\beta) - \rho(e_i - x_i'\beta^*) \\
&\quad = \psi(e_i - x_i'\beta)x_i'(\beta^* - \beta) + \psi'(\theta_1)(x_i'(\beta^* - \beta))^2/2 \\
&\quad = \psi(e_i)x_i'(\beta^* - \beta) - \psi'(\theta_2)x_i'\beta x_i'(\beta^* - \beta) \\
&\qquad + \psi'(\theta_1)(x_i'(\beta^* - \beta))^2/2.
\end{aligned}$$

式中 θ_1 介于 $e_i - x_i'\beta$ 和 $e_i - x_i'\beta^*$ 之间;而 θ_2 介于 e_i 与 $e_i - x_i'\beta$ 之间,由此知存在常数 M 充分大,使当 n 充分大时,有

$$\begin{aligned}
A_n(\beta) | &\equiv \left| \sum_{i=1}^{n}(\rho(e_i - x_i'\beta) - \rho(e_i - x_i'\beta^*)) \right| \\
&\leqslant M\left(\sum_{i=1}^{n} |\psi(e_i)| n^{-3} + n^{-2} + n^{-5} \right).
\end{aligned} \tag{6.73}$$

这里用到了

$$\begin{aligned}
\sum_{i=1}^{n}(x_i'(\beta - \beta^*))^2 &= (\beta - \beta^*)' S_n (\beta - \beta^*) \\
&\leqslant \mu_n \| \beta - \beta^* \|^2 \leqslant p\mu_n n^{-6},
\end{aligned}$$

又因为 $\{x_i\}$ 有界,有 $\mu_n = O(n)$,故

$$\sum_{i=1}^{n}(x_i'(\beta - \beta^*))^2 = O(n^{-5}).$$

因为
$$P\Big(\sum_{i=1}^{n}|\psi(e_i)|n^{-3}\geqslant 1\Big)\leqslant n^{-3}\sum_{i=1}^{n}E|\psi(e_i)|$$
$$=O(n^{-2}),$$
由此及(6.73)知：若取常数 M_1 充分大，有
$$P\Big(\sup_{\beta\in B_n}|A_n(\beta)|\geqslant M_1\Big)=O(n^{-2}).\qquad(6.74)$$
定义事件
$$E_n=\{\inf_{\beta\in B_n}\Delta_n(\beta)\leqslant 0\},$$
因为
$$\Delta_n(\beta)=A_n(\beta)+\Delta_n(\beta^*),$$
有
$$W_n\equiv\inf_{\beta\in B_n}\Delta_n(\beta)\geqslant-\sup_{\beta\in B_n}|A_n(\beta)|+\inf_{\beta\in G_n}\Delta_n(\beta).$$
由此式及(6.74)，得
$$P(W_n\leqslant-M_1+\inf_{\beta\in G_n}\Delta_n(\beta))=O(n^{-2}).\qquad(6.75)$$
由 $\hat{\beta}_n$ 的定义，有
$$E_n\supset\{\|S_n^{1/2}\hat{\beta}_n\|\geqslant n^{\alpha}\},$$
且当 n 充分大至使 $2^{-1}h_1n^{2\alpha}>M_1$ 时，有
$$E_n\subset\Big\{W_n\leqslant-M_1+\inf_{\beta\in G_n}\Delta_n(\beta)\Big\}\cup\Big\{\inf_{\beta\in G_n}\Delta_n(\beta)\leqslant 2^{-1}h_1n^{2\alpha}\Big\}.\qquad(6.76)$$

现取 $m\geqslant 2$ 充分大，使 $2\alpha m-3p-2\geqslant 2$，则由(6.72)，有
$$P\Big(\inf_{\beta\in G_n}\Delta_n(\beta)\leqslant 2^{-1}h_1n^{2\alpha}\Big)=O(n^{-2}).$$
由这个事实以及(6.75)和(6.76)，得 $P(E_n)=O(n^{-2})$. 于是由 Borel-Cantelli 引理，得 $P(E_n,\ \text{i.o.})=0$. 再由事件
$$E_n\supset\{\|S_n^{1/2}\hat{\beta}_n\|\geqslant n^{\alpha}\},$$
即得(6.69). 引理证毕. ▌

三、定理 6.7 的证明

不失普遍性，仍设 $\beta_0=0$. 令
$$\beta_n^*=S_n^{1/2}\hat{\beta}_n,\ \overline{\beta}_n=g^{-1}\sum_{i=1}^{n}\psi(e_i)x_{ni},\qquad(6.77)$$

$$H_n = \sum_{i=1}^{n}(\rho(e_i - x_{ni}'\beta_n^*) - \rho(e_i - x_{ni}'\overline{\beta}_n)). \qquad (6.78)$$

按 $\hat{\beta}_n$ 的定义，有 $H_n \leqslant 0$．把 $\rho(e_i - x_{ni}'\beta_n^*)$ 在点 $e_i - x_{ni}'\overline{\beta}_n$ 附近作 Taylor 展开，有

$$\begin{aligned}
H_n = & \sum_{i=1}^{n}\psi(e_i - x_{ni}'\overline{\beta}_n)x_{ni}'(\overline{\beta}_n - \beta_n^*) \\
& + 2^{-1}(\beta_n^* - \overline{\beta}_n)'\sum_{i=1}^{n}\psi'(e_i - x_{ni}'\overline{\beta}_n \\
& + \theta_{1i}x_{ni}'(\overline{\beta}_n - \beta_n^*))x_{ni}x_{ni}'(\beta_n^* - \overline{\beta}_n) \\
= & \sum_{i=1}^{n}\psi(e_i)x_{ni}'(\overline{\beta}_n - \beta_n^*) - \overline{\beta}_n\sum_{i=1}^{n}\psi'(e_i)x_{ni}x_{ni}'(\overline{\beta}_n - \beta_n^*) \\
& - \overline{\beta}_n\sum_{i=1}^{n}(\psi'(e_i - \theta_{2i}x_{ni}'\overline{\beta}_n) - \psi'(e_i))x_{ni}x_{ni}'(\overline{\beta}_n - \beta_n^*) \\
& + 2^{-1}g\|\beta_n^* - \overline{\beta}_n\|^2 \\
& + 2^{-1}(\beta_n^* - \overline{\beta}_n)'\sum_{i=1}^{n}(\psi'(e_i) - g)x_{ni}x_{ni}'(\beta_n^* - \overline{\beta}_n) \\
& + 2^{-1}(\beta_n^* - \overline{\beta}_n)'\sum_{i=1}^{n}(\psi'(e_i - x_{ni}'\overline{\beta}_n + \theta_{1i}x_{ni}'(\overline{\beta}_n - \beta_n^*)) \\
& - \psi'(e_i))x_{ni}x_{ni}'(\beta_n^* - \overline{\beta}_n).
\end{aligned}$$

上式中 $|\theta_{ji}| \leqslant 1$ $(j = 1,2; i \geqslant 1)$，定义 A_n 如(6.62)．注意到 $\overline{\beta}_n$的定义(6.77)．易见上式右端前两项之和等于 $J_1 \equiv -\overline{\beta}_n'A_n(\overline{\beta}_n - \beta_n^*)$．把上式右端第 3、5、6 项分别记为 J_2、J_3、J_4，得

$$H_n = 2^{-1}g\|\beta_n^* - \overline{\beta}_n\|^2 + \sum_{i=1}^{4}J_i. \qquad (6.79)$$

注意到

$$|\theta_{ji}| \leqslant 1,\ \sum_{i=1}^{n}\|x_{ni}\|^2 = p$$

及 ψ' 满足 Lip(1)，有

$$|J_1| \leqslant C|A_n| \cdot \|\overline{\beta}_n\| \cdot \|\beta_n^* - \overline{\beta}_n\|,$$

$$|J_2| \leqslant C\sqrt{d_n}\|\beta_n\|^2 \cdot \|\beta_n^* - \overline{\beta}_n\|,$$

$$|J_3|\leqslant C|A_n|\cdot\|\beta_n^*-\overline{\beta}_n\|^2,$$

$$|J_4|\leqslant C(\sqrt{d_n}\,\|\overline{\beta}_n\|+\sqrt{d_n}\,\|\beta_n^*\|)\|\beta_n^*-\overline{\beta}_n\|^2.$$

这里 C 为常数，与 n 无关. 又注意到 $|A_n|$ 表示方阵 A_n 的各元绝对值的最大值，由以上估计及(6.79)式，有

$$\begin{aligned}0\geqslant H_n\geqslant \|\beta_n^*-\overline{\beta}_n\|\{(g/2-C|A_n|\\ -C\sqrt{d_n}\,\|\beta_n^*\|-C\sqrt{d_n}\,\|\overline{\beta}_n\|)\|\beta_n^*-\overline{\beta}_n\|\\ -C|A_n|\cdot\|\overline{\beta}_n\|-C\sqrt{d_n}\,\|\overline{\beta}_n\|^2\}.\end{aligned}\tag{6.80}$$

按(6.63)及假定 $d_n=O(n^{-\delta})$，有 $|A_n|=o(1)$，a.s.，又据引理6.10，知(6.48)成立，于是有

$$\sqrt{d_n}\,\|\overline{\beta}_n\|=o(1),\ \text{a.s.}.$$

在引理 6.12 中取 $\alpha<\delta/2$，则由引理 6.12 及 $d_n=O(n^{-\delta})$，有

$$\sqrt{d_n}\,\|\beta_n^*\|=o(1),\ \text{a.s.}.$$

又

$$\begin{aligned}|A_n|\cdot\|\overline{\beta}_n\|&=O(\sqrt{d_n}\log n),\ \text{a.s.},\\ \sqrt{d_n}\,\|\overline{\beta}_n\|^2&=O(\sqrt{d_n}\log n),\ \text{a.s.}.\end{aligned}\tag{6.81}$$

把以上的估计用于(6.80)，得

$$\begin{aligned}0\geqslant H_n\geqslant \|\beta_n^*-\overline{\beta}_n\|\{(g/2-o(1))\|\beta_n^*-\overline{\beta}_n\|\\ -O(\sqrt{d_n}\log n)\},\text{a.s.}.\end{aligned}\tag{6.82}$$

因为 $g/2>0$，由(6.81)即得(6.52).

如果 $\lambda_n\sim\mu_n$，则根据引理 6.10 及(6.64)，(6.81)中的 $\log n$ 可代之以 $\log\log n$，因而(6.82)式中的 $\log n$ 也可代之以 $\log\log n$. 这导致(6.54). 以上证明了定理 6.7 的强表示部分.

下面转到证明定理 6.7 的弱表示部分：注意到

$$\|\overline{\beta}_n\|=O_p(1),\quad |A_n|=O_p(\sqrt{d_n}).\tag{6.83}$$

其证明是容易的：

$$E\|\overline{\beta}_n\|^2=g^{-2}\sum_{i=1}^n\|x_{ni}\|^2E(\psi'(e))^2=g^{-2}p\sigma^2$$

与 n 无关. 按引理 6.11 的证明,取 A_n 的一个元如(6.65)式中的 η_n,根据 ψ' 的有界性,易得

$$E\eta_n^2 = O(d_n).$$

由此推出(6.83). 将(6.83)用于(6.80),得

$$0 \geqslant H_n \geqslant \| \beta_n^* - \bar{\beta}_n \| \{ (g/2 - o_p(1)) \| \beta_n^* - \bar{\beta}_n \| - O_p(\sqrt{d_n}) \}.$$

由此式即得(6.53).

最后要证明(6.53)不能一般地被改善. 为此,取位置参数模型 $Y_i = \beta_0 - e_i (1 \leqslant i \leqslant n, n \geqslant 1)$,有 $d_n = n^{-1}$. 选定 ρ 及 e_1 的分布,使之满足定理 6.7 的各条件,且 ψ'' 和 ψ''' 在 R^1 处处存在并有界,$E\psi''(e_1) = 0$. 注意到这样一个实例:

$$\rho(u) = u^4/(1+u^2)^2, \quad e_1 \text{ 的分布为 } N(0,1).$$

我们来验证:此例满足定理 6.7 的一切条件,只有 4°需要讨论. 其他条件都是显然的,有

$$h(t) \equiv E\rho(e_1 + t) = \int_{-\infty}^{\infty} \exp(-(u-t)^2/2) u^4 (1+u^2)^{-1} du,$$

而

$$h'(t) = \int_{-\infty}^{\infty} (u-t) \exp(-(u-t)^2/2) u^4 (1+u^2)^{-1} du$$
$$= \int_0^{\infty} u e^{-u^2/2} \left(\frac{(u+t)^4}{1+(u+t)^2} - \frac{(u-t)^4}{1+(u-t)^2} \right) du.$$

易见函数 $u^4(1+u^2)^{-1}$ 当 $u > 0$ 时是严格增的,故由上面后一等式知当 $t > 0$ 时,$h'(t) > 0$;另一方面,由上面前一等式,知 $h'(-t) = -h'(t)$,故知当 $t < 0$ 时,$h'(t) < 0$. 这样一来,$h(t)$ 在 $t < 0$ 处严格下降,而在 $t > 0$ 处则严格上升. 因而必以 $t = 0$ 为其唯一的最小值点.

以下仍设 $\beta_0 = 0$. β_0 的 M 估计 $\hat{\beta}_n$ 为方程

$$\sum_{i=1}^{n} \psi(e_i - \beta) = 0$$

的解,其中 $\psi = \rho'$. 作 Taylor 展开,可以把

$$\sum_{i=1}^{n}\psi(e_i-\hat{\beta}_n)=0$$

写为

$$\sum_{i=1}^{n}\psi(e_i)-\sum_{i=1}^{n}\psi'(e_i)\hat{\beta}_n+2^{-1}\sum_{i=1}^{n}\psi''(e_i-\theta_i\hat{\beta}_n)\hat{\beta}_n^2$$
$$=0\quad(0\leqslant\theta_i\leqslant 1).\tag{6.84}$$

因为对本例而言有 $\overline{\beta}_n=O_p(1)$ ，故由(6.53)推出

$$\hat{\beta}_n=O_p(n^{-1/2})=o_p(1).$$

于是，因为 ψ'' 有界，有

$$\psi''(e_i-\theta_i\hat{\beta}_n)=\psi''(e_i)+O_p(1)\hat{\beta}_n,\tag{6.85}$$

以此代入(6.84)，并注意

$$\sum_{i=1}^{n}\psi(e_i)-ng\hat{\beta}_n=n^{1/2}g(\overline{\beta}_n-n^{1/2}\hat{\beta}_n)=-n^{1/2}gR_n,$$

有

$$0=-n^{1/2}gR_n-\sum_{i=1}^{n}(\psi'(e_i)-g)\hat{\beta}_n+2^{-1}\sum_{i=1}^{n}\psi''(e_i)\hat{\beta}_n^2+O(1)n\hat{\beta}_n^3$$
$$\equiv -n^{1/2}gR_n-L_1+L_2+L_3.\tag{6.86}$$

因为 ψ'' 有界，有

$$\sum_{i=1}^{n}\psi''(e_i)=O_p(n^{1/2}).$$

又

$$\hat{\beta}_n=O_p(n^{-1/2}),$$

故有

$$L_2=O_p(n^{-1/2})=o_p(1).$$

因此，若

$$R_n=o_p(d_n^{1/2})=o_p(n^{-1/2}),$$

则由(6.86)将得到

$$L_1=o_p(1),$$

但这是不可能的，因为由

$$\sigma^2\equiv E(\psi'(e_1))^2>0$$

知

$$\sum_{i=1}^{n}(\psi'(e_i)-g)/\sqrt{n}\xrightarrow{\mathscr{L}}N(0,\sigma^2),$$

而

$$\sqrt{n}\beta_n\xrightarrow{\mathscr{L}}N(0,\sigma^2/g^2),$$

故

$$L_1=\sqrt{n}\beta_n\cdot\sum_{i=1}^{n}(\psi'(e_i)-g)/\sqrt{n}$$

不可能依概率收敛于0.这证明了R_n不可能为$o_p(n^{-1/2})$.于是定理6.7全部证毕. ▎

注1 由于弱表示(6.53)中的$\sqrt{d_n}$部分已一般地不可改善,因此,对强表示(6.52)和(6.54),其中的$\sqrt{d_n}$部分当然也无可改善,$\log n$或$\log\log n$部分另当别论.

注2 从本例讨论过程中可看出,(6.53)不可改进,根源在于$P(\psi'(e_1)=g)\neq 1$,而这是一般的情况.如果$P(\psi'(e_1)=g)=1$,则(6.53)确可以改进.一个平凡的例子是$\rho(u)=u^2$ (LS估计).这时$R_n\equiv 0$.也可以举出非LS估计的例子来.

例6.6 仍考虑位置参数模型$Y_i=\beta_0+e_i$.在此取$\rho(u)=u^2/(1+u^2)$,e_1的分布为

$$P(e_1=a)=P(e_1=-a)=1/2,\ 0<a<(2\sqrt{3})^{-1}.$$

容易验证,这满足定理6.7的所有条件,只有条件4°的验证较复杂一点(此例中ρ为非凸函数,故这一条并非自明),有

$$E\rho(e_1+t)=\frac{1}{2}\left(\frac{(a+t)^2}{1+(a+t)^2}+\frac{(a-t)^2}{1+(a-t)^2}\right),$$

为证$t=0$是$E\rho(e_1+t)$的唯一最小值点,只须证明:函数

$$m(t)=(1+(t+a)^2)^{-1}+(1+(t-a)^2)^{-1}$$

有唯一最大值点$t=0$.因为当$t\geqslant a$时,$m(t)\leqslant m(a)$.故只须考虑$0\leqslant t\leqslant a$的情况.记

$$f(t)=2t(1+t^2)^{-2},$$

$$f'(t)=(2-6t^2)(1+t^2)^{-3},$$

则 $$m'(t) = -f(a+t) + f(a-t). \tag{6.87}$$
因为当 $|t| < 1/\sqrt{3}$ 时 $f'(t) > 0$，知在这个范围内 f 严格增加. 当 $0 \leqslant t \leqslant a < 1/(2\sqrt{3})$ 时，有 $|a+t| < 1/\sqrt{3}$，$|a-t| < 1/\sqrt{3}$. 故当 $0 < a < 1/(2\sqrt{3})$，而 t 由 0 增加至 a 时，(6.87)右端两项都严格下降. 又 $m'(0) = 0$，故当 $0 < t \leqslant a$ 时 $m'(t) < 0$，即 $m(t)$ 在 $0 \leqslant t \leqslant a$ 内严格下降，而 0 为 $m(t)$ 的唯一最大值点. 对本例而言，(6.86)式中的 L_1 为 0，而 L_2 和 L_3 则都是 $O_p(n^{-1/2})$，故
$$R_n = O_p(n^{-1}) = O_p(d_n),$$
超过了(6.53)给出的阶 $O_p(\sqrt{d_n})$. 由(6.86)也可以看出，除非 $\rho(u) = u^2$，不能有比 $O_p(n^{-1})$ 更高的阶.

在本例中，强表示(6.54)也可以改善：应用重对数律，并注意到(6.86)式中的 L_1 为 0，易得
$$R_n = O(n^{-1}(\log\log n)^{3/2}) = O(d_n(\log\log n)^{3/2}),\ \text{a.s.}.$$

§6.4　线性表示的应用

线性表示(6.2)把 M 估计分解为一个线性独立和与一个剩余项，而后者很小，起主要作用的乃是前者. 利用这一点，可得出有关 M 估计的若干深层次的渐近性质. 本节中将讨论这类几个问题，以印证上面的论断.

一、*M* 估计强收敛速度

在第 3 章中曾证明了：在很一般的条件下，M 估计 $\hat{\beta}_n$ 以概率 1 收敛到参数真值 β_0. 这收敛的速度有多快? 在第 3 章中未曾回答. 应用线性表示，容易证明下面的结果：

定理 6.8　在定理 6.3 或定理 6.7 的条件 1°～5°之下，有
$$\hat{\beta}_n - \beta_0 = O((d_n \log n)^{1/2}),\ \text{a.s.}. \tag{6.88}$$
而在定理 6.1 或定理 6.7 的条件 1°～5°加上 $\lambda_n \sim \mu_n$ 之后，(6.88)中的 $(\log n)^{1/2}$ 可改进为 $(\log\log n)^{1/2}$.

证　有

$$\hat{\beta}_n - \beta_0 = S_n^{-1/2}\bar{\beta}_n + S_n^{-1/2}R_n,$$

式中 $\bar{\beta}_n$ 系按(6.77)的定义. 因为

$$|S_n^{-1/2}| = O(\lambda_n^{-1/2}),$$

故根据引理 6.8,有

$$|S_n^{-1/2}| = O(d_n^{1/2})$$

(注意:引理 6.8 的这一部分并不要求 $\{x_i\}$ 有界). 因此,不论在定理 6.1、定理 6.3 或定理 6.5 之下,都有

$$S_n^{-1/2}R_n = o(d_n^{1/2}) = o((d_n\log n)^{1/2}), \text{ a.s.}.$$

而根据引理 6.3 或引理 6.7 或引理 6.10,都有

$$\bar{\beta}_n = O((\log n)^{1/2}), \text{ a.s.},$$

因而

$$S_n^{-1/2}\bar{\beta}_n = O((d_n\log n)^{1/2}), \text{ a.s.}.$$

将二者结合起来,即得(6.88). 若进一步假定 $\lambda_n \sim \mu_n$,则

$$\bar{\beta}_n = O((\log n)^{1/2}), \text{ a.s.},$$

可以用

$$\bar{\beta}_n = O((\log\log n)^{1/2}), \text{ a.s.}$$

代替之,因而(6.88)中的 $(\log n)^{1/2}$ 可以用 $(\log\log n)^{1/2}$ 代替. 定理证毕. ▌

二、M 估计的重对数律

以下分别用 β_{0j}, $\hat{\beta}_{nj}$, $\bar{\beta}_{nj}$ 和 ξ_{nj} 记 β_0, $\hat{\beta}_n$, $\bar{\beta}_n$ 和 $S_n^{1/2}(\hat{\beta}_n - \beta_0)$ 的第 j 个分量 $(1 \leqslant j \leqslant p)$.

定理 6.9 设定理 6.1 或定理 6.7 的条件满足,且存在常数列 $\{c_n\}$ 及正定方阵 V,使

$$S_n/c_n \to V, \tag{6.89}$$

则有

$$\lim{}^{\sup}_{\inf}\left(\xi_{nj}/(2g^{-2}\sigma^2\log\log c_n)^{1/2}\right) = \begin{cases} 1 \\ -1 \end{cases}, \text{ a.s.}, \tag{6.90}$$

$$\lim{}^{\sup}_{\inf}\left((\hat{\beta}_{nj} - \beta_{0j})/(2c_n^{-1}v_j^{(-1)}g^{-2}\sigma^2\log\log c_n)^{1/2}\right) = \begin{cases} 1 \\ -1 \end{cases}, \text{ a.s.}. \tag{6.91}$$

σ^2 和 g 的意义见定理 6.1 的条件 A_8 和 A_9，$v_j^{(-1)}$ 是方阵 V^{-1} 的 (j,j) 元. 又在定理 6.7 的条件下，以上两式中的 $\log\log c_n$ 可改为 $\log\log n$.

证　令

$$T_n = \sum_{i=1}^{n} V^{-1/2} x_i \psi(e_i)$$

并以 a_{ij} 记 $V^{-1/2}x_i$ 的第 j 个分量，则

$$T_{nj} \equiv T_n \text{ 的第 } j \text{ 个分量} = \sum_{i=1}^{n} a_{ij}\psi(e_i).$$

不难验证，在本定理的条件下，引理 6.3 中对 Q_n 所作的论证完全适用于此处的 T_{nj}. 由于

$$\mathrm{Var}(T_{nj}) = \sigma^2 \sum_{i=1}^{n} a_{ij}^2 = \sigma^2 \cdot V^{-1/2} S_n V^{-1/2} \text{ 的}(j,j)\text{ 元},$$

而根据(6.89)，有

$$V^{-1/2} S_n V^{-1/2} = c_n I_p + o(c_n),$$

于是

$$\sum_{i=1}^{n} a_{ij}^2 / c_n \to 1.$$

这样，类似于(6.17)，得到

$$\lim_{\inf}^{\sup} (T_{nj}/(2c_n\sigma^2 \log\log c_n)^{1/2}) = \begin{cases} 1 \\ -1 \end{cases}, \text{ a.s.}. \qquad (6.92)$$

再利用(6.89)，得到

$$\bar{\beta}_n = g^{-1} S_n^{-1/2} V^{1/2} T_n = g^{-1} c_n^{-1/2} T_n + o(c_n^{-1/2} T_n). \quad (6.93)$$

由(6.92)与(6.93)，得

$$\lim_{\inf}^{\sup} (\bar{\beta}_{nj}/(2g^{-2}\sigma^2 \log\log c_n)^{1/2}) = \begin{cases} 1 \\ -1 \end{cases}, \text{ a.s.}, \quad (6.94)$$

在本定理的条件下，根据定理 6.1 或定理 6.7 的结论以及 $c_n \to \infty$（这由 $d_n \to 0$，因而 $S_n^{-1} \to 0$ 以及(6.89)推出），由(6.94)即得(6.90).

(6.91)的证明类似，只须把 T_n 的定义修改为

$$T_n = \sum_{i=1}^{n} V^{-1} x_i \psi(e_i),$$

并以 a_{ij} 记 $V^{-1}x_i$ 的第 j 个分量，注意

$$\begin{aligned}\sum_{i=1}^{n} a_{ij}^2 &= V^{-1}S_nV^{-1} \text{ 的}(j,j)\text{ 元} \\ &= V^{-1/2}(V^{-1/2}S_nV^{-1/2})V^{-1/2} \text{ 的}(j,j)\text{ 元} \\ &= V^{-1/2}(c_nI_p + o(c_n))V^{-1/2} \text{ 的}(j,j)\text{ 元} \\ &= (c_nV^{-1} + o(c_n)) \text{ 的}(j,j)\text{ 元} \\ &= c_nv_j^{-1}(1 + o(1)).\end{aligned}$$

因而

$$\sum_{i=1}^{n} a_{ij}^2/(c_n v_j^{(-1)}) \to 1.$$

其次

$$gS_n^{-1/2}\bar{\beta}_n = S_n^{-1/2}S_n^{-1/2}VT_n = S_n^{-1}VT_n = c_n^{-1}T_n + o(c_n^{-1}T_n),$$

其余的推导与前无异.

最后，在定理 6.7 的条件下，根据假定 $d_n = O(n^{-\delta})$ 以及引理 6.8，有

$$n^{\delta} = O(d_n^{-1}) = O(\lambda_n).$$

而根据(6.89)，有 $\lambda_n = O(c_n)$，故 $n^{\delta} = O(c_n)$. 另一方面，根据假定 $\{x_i\}$ 有界，又有 $c_n = O(n)$. 将二者结合，得

$$\log\log c_n/\log\log n \to 1.$$

因而在(6.90)和(6.91)两式中，$\log\log c_n$ 可改为 $\log\log n$. 至此完成了本定理的证明. ▌

三、M 估计的分布收敛于正态分布的一致速度

ξ_{nj} 的意义与第二段中相同. 以 F_n 记 $g\xi_{nj}/\sigma$ 的分布函数；以 Φ 记 $N(0,1)$ 的分布函数.

定理 6.10 设定理 6.7 的条件 1°～5°成立，则

$$\sup_x |F_n(x) - \Phi(x)| = O(d_n^{1/2}\log n). \tag{6.95}$$

证 仍如前设 $\beta_0 = 0$，β_n^* 和 $\bar{\beta}_n$ 的定义如(6.77)，A_n 的定义如(6.62). 定义事件

$$E_n = \{\|\beta_n^*\| \leqslant n^\alpha\} \cap \{|A_n| < \epsilon\} \cap \{\sqrt{d_n}\,\|\bar{\beta}_n\| < \epsilon\}, \tag{6.96}$$

式中 $0 < \epsilon < (8C)^{-1}g$，$C$ 的意义见(6.80)，而 $\alpha = \delta/4$．取 A_n 的 (j,k) 元并记之为 T_n．以 x_{nij} 记 x_{ni} 的第 j 个分量，则因

$$\sum_{i=1}^n x_{nij}^2 x_{nik}^2 \leqslant \max_{1\leqslant i\leqslant n} x_{nij}^2 \sum_{i=1}^n x_{nik}^2 \leqslant d_n = O(n^{-\delta}),$$

由此，注意到 ψ' 有界，因而 $\psi'(e_i) - g$ 的任意阶矩存在，引用(6.68)，得知对不小于 $2/\delta$ 的整数 r，有

$$ET_n^{2r} = O(n^{-r\delta}) = O(n^{-2}),$$

故

$$P(|T_n| \geqslant \epsilon) = O(n^{-2}).$$

用完全同样的方法，可证明

$$P(\sqrt{d_n}\,\|\beta_n\| \geqslant \epsilon) = O(n^{-2}).$$

又根据引理 6.12 的证明的结尾处知

$$P(\|\beta_n^*\| \geqslant n^\alpha) = O(n^{-2}).$$

总结以上，得到

$$P(E_n) \geqslant 1 - O(n^{-2}). \tag{6.97}$$

于是，根据

$$0 < \epsilon < (8C)^{-1}g, \quad \alpha = \delta/4, \quad d_n = O(n^{-\delta}),$$

知当事件 E_n 发生且 n 充分大时，有

$$g/2 - C|A_n| - C\sqrt{d_n}\,\|\beta_n^*\| - C\sqrt{d_n}\,\|\bar{\beta}_n\| \geqslant g/4.$$

将上式与(6.80)结合，可知当 E_n 发生且 n 充分大时，有

$$\|\beta_n^* - \bar{\beta}_n\| \leqslant C_0|A_n| \cdot \|\bar{\beta}_n\| + C_0\sqrt{d_n}\,\|\beta_n\|^2. \tag{6.98}$$

式中 C_0 为某一常数．以 $\bar{F}_n$ 记 $g\beta_{nj}/\sigma$ 的分布，则据引理 6.1，有

$$\sup_x |\bar{F}_n(x) - \Phi(x)| = O(\sqrt{d_n}). \tag{6.99}$$

定义事件

$$E_{n1j} = \{|\bar{\beta}_{nj}| \leqslant \sigma\sqrt{\log n}/g\},$$

并记

$$E_{n1}=\bigcap_{j=1}^{p}E_{n1j}.$$

则根据(6.99),有

$$\begin{aligned}P(E_{n1j}^c)&\leqslant O(\sqrt{d_n})+2(1-\Phi(\sqrt{\log n}))\\&=O(\sqrt{d_n})+O(n^{-1/2})\\&=O(\sqrt{d_n}).\end{aligned}\tag{6.100}$$

记　　$$\sigma_0^2=\mathrm{Var}(\psi'(e_1)),$$

定义事件

$$E_{n2jk}=\{|T_n|\leqslant(3\sigma_0^2d_n\log n)^{1/2}\},\quad E_{n2}=\bigcap_{j,k=1}^{p}E_{n2jk}.$$

式中如前 T_n 是 A_n 的 (j,k) 元. 由于 ψ' 有界,对某个常数 b,有

$$\sup|x_{nij}x_{nik}(\psi'(e_i)-g)|\leqslant bd_n\quad(1\leqslant i\leqslant n).$$

又

$$\mathrm{Var}(T_n)\leqslant d_n\sigma_0^2.$$

由此,利用引理 3.3,得到

$$P(E_{n2jk}^c)\leqslant 2\exp\left(-\frac{3\sigma_0^2d_n\log n}{2d_n\sigma_0^2+2bd_n(3\sigma_0^2d_n\log n)^{1/2}}\right),\tag{6.101}$$

由于 $d_n=O(n^{-\delta})$,知 $d_n\log n\to 0$. 故当 n 充分大时,有

$$2b(3\sigma_0^2d_n\log n)^{1/2}\leqslant\sigma_0^2.$$

所以,由(6.101)即知,当 n 充分大时,有

$$P(E_{n2jk}^c)\leqslant 2\exp(-\log n)=2/n=O(d_n).\tag{6.102}$$

结合(6.97)、(6.100)和(6.102),得

$$P(E_n\cap E_{n1}\cap E_{n2})\geqslant 1-O(\sqrt{d_n}).\tag{6.103}$$

但由(6.98),及事件 E_{n1} 和 E_{n2} 的定义可知,当 $E_n\cap E_{n1}\cap E_{n2}$ 发生并且 n 充分大时,有

$$g|\beta_{nj}-\xi_{nj}|/\sigma\leqslant C_1\sqrt{d_n}\log n.\tag{6.104}$$

式中 C_1 为某一常数(注意,按我们的记号,ξ_{nj} 为 β_n^* 的第 j 个分量). 由此,利用(6.99),并注意 Φ' 在 R^1 上有界,得

$$F_n(x)=P(g\xi_{nj}/\sigma\leqslant x)\leqslant P(g\overline{\beta}_{nj}/\sigma\leqslant x+C_1\sqrt{d_n}\log n)+O(\sqrt{d_n})$$
$$=\overline{F}_n(x+C_1\sqrt{d_n}\log n)+O(\sqrt{d_n})$$
$$=\overline{F}_n(x+C_1\sqrt{d_n}\log n)-\Phi(x+C_1\sqrt{d_n}\log n)$$
$$+(\Phi(x+C_1\sqrt{d_n}\log n)-\Phi(x))+\Phi(x)+O(\sqrt{d_n})$$
$$=O(\sqrt{d_n})+O(\sqrt{d_n}\log n)+\Phi(x)+O(\sqrt{d_n})$$
$$=\Phi(x)+O(\sqrt{d_n}\log n). \tag{6.105}$$

同理,从

$$F_n(x)\geqslant\overline{F}_n(x-C_1\sqrt{d_n}\log n)-O(\sqrt{d_n})$$

出发,得到

$$F_n(x)\geqslant\Phi(x)+O(\sqrt{d_n}\log n). \tag{6.106}$$

将(6.105)与(6.106)结合起来,即得(6.95).定理证毕. ∎

关于(6.95)的精度,可以指出的是:能构造出这样的例子:使(6.95)的左端达不到 $o(\sqrt{d_n})$;而至今尚未造出这样的例子:使(6.95)的左端不是 $O(\sqrt{d_n})$.因此,有这种可能:(6.95)可改进为 $O(\sqrt{d_n})$.这一点是否正确尚待研究.如果正确,则 $O(\sqrt{d_n})$ 就是可能达到的最好的结果.

参 考 文 献

[1] 陈希孺. 线性估计弱相合的一个问题. 数学年刊,1981,(2):131~137

[2] 陈希孺. 线性模型参数 M 估计的线性表示. 中国科学 A 辑,1993,(23): 1264~1275

[3] 陈希孺,方兆本,李国英,陶波. 非参数统计. 上海科学技术出版社,1989

[4] 陈希孺,白志东,赵林城,吴月华. 线性模型中最小二乘估计的渐近正态性. 中国科学 A 辑, 1990,(20):449~463

[5] 陈希孺,陈桂景,吴启光,赵林城. 线性模型参数的估计理论. 北京:科学出版社, 1985

[6] 陶波,成平. 关于矩的一个不等式. 数学年刊,1981,(2):451~461

[7] Amemiya T. Two stage least absolute deviation estimators, Econometrika,1982, (50): 689~711

[8] Anderson P K and Gill R D. Cox's regression model for counting processes: A large sample study. Ann Statist,1982,(10): 1100~1120

[9] Babu G J. Strong representations for LAD estimates in linear models. Probab. Theory Related Fields, 1989,(83F):547~558

[10] Bahadur R R. A note on quantiles in large samples, Ann Math Statist,1966, (37): 577~580

[11] Bai Z D, Chen X R, Miao B Q and Rao C R. Asymptotic theory of least distances estimate in multivariate linear models. Statistics,1990,(21):503~519

[12] Bai Z D, Rao C R and Wu Y. M-estimation of multivariate linear regression parameters under a convex discrepancy function,Statistica Sinica,1992,(2):237~254

[13] Bai Z D, Rao C R and Yin Y Q. Least absolute deviations analysis of variance, Sanhkyā Ser A,1990,(52):166~177

[14] Bai Z D, Rao C R and Zhao L C. MANOVA Type tests under a convex discrepancy function for the standard multivariate linear model, J statist planning and Inference, 1993,(36):77~90

[15] Bai Z D,Rao C R and Zhao L C. Weak representation of regression estimates in multivariate linear models obtained by minimizing a convex function of residuals. Tech Report Penn State Univ, 1992

[16] Bai Z D, Wu Y H, Chen X R and Miao B Q. On solvability of an equation arising in the theory of M-estimates, Comm Statist Theory and Methods, 1990,(19):

363～380

[17] Bassett G and Koenker R. Asymptotic theory of least absolute error regression, J Amer Statist Assoc, 1978, (73): 618～622

[18] Bennett G. Probability inequalities for the sums of independent random variables. J Amer Statist Assoc, 1962, (57): 33～45

[19] Bickel P J. One step Huber estimates in the linear model. J Amer Statist Assoc, 1975, (70): 428～433

[20] Bloomfield P and Steiger W L. Least Absolute Deviations. Birkhauser, 1983

[21] Chen Xiru. Consistency of least squares estimates in linear models. Scientia Sinica Special Issue (Ⅱ), 1979, 162～176

[22] Chen Xiru, Bai Zhidong, Zhao Lincheng and Wu Yuehua. Consistency of Minimum L_1-Norm estimates in linear models, in The development of statistics: Recent contributions from China. Pitman Research Notes in Mathematics Series 258, Longman Scientific & Technical, 1992, 249～260

[23] Chen X R and Wu Y H. Strong consistency of M-estimates in linear models. J Multivariate Anal, 1988, (27): 116～130

[24] Chen X R and Wu Y H. On a necessary condition for the consistency of the L_1 estimates in linear models. Comm Statist, Theory and Metheds, 1993, (22): 631～639

[25] Chen X R, Zhao L C and Wu Y H. On conditions of consistency of ML_1N estimates. Statistica Sinica, 1993, (3): 9～18

[26] Dodge Y, Jureckova J. Adaptive combination of least squares and least absolute deviations estimates. In Statistical Data Analysis Based on the L_1-norm and Related Methods. North Holland, 1987, 275～284

[27] Drygas H. Weak and strong consistency of the least squares estimates in regression model. Z Wahrsch. Verw Gebiete, 1976, (34): 119～127

[28] Dupacŏvá J. Asymptotic properties of restricted L_1-estimates of regression. In Statistical Data Analysis Based on the L_1-Norm and Related Methods. North-Holland, 1987, 263～274

[29] Fuk D Hk and Nagaev S V. Probability inequalities for sums of independent random variables. Theory Probab. Appl, 1971, (16): 643～660

[30] Heiler S and Willers R. Asymptotic normality of R-estimates in the linear model. Statistics, 1988, (19): 173～184

[31] Hoeffding W. Probability inequalities for sums of bounded random variables. J Amer Statist Assoc, 1963, (58): 13～30

[32] Hsu P L and Robbins H. Complete convergence and the law of large numbers.

Proc Nat Acad Sci U S A,1947,(33):25～31

[33] Huber P J. Robust estimation of a location parameter. Ann Math Statist, 1964, (35):73～101

[34] Huber P J. Robust regression. Ann Statist,1973,(1):799～821

[35] Huber P J. The place of the L_1-norm in robust estimation. In Statistical Data Analysis Based on the L_1-Norm and Related Methods. North-Holland. 1987

[36] Humark K M S. Statistische Methoden der Modelbildung Ⅰ. Akademie-Velag, 1983

[37] Jackel L A. Estimating regression coefficients by minimizing the dispersion of the residuals. Ann Math Statist,1972,(43):1449～1459

[38] Jurečková J. Nonparametric estimate of regression coefficients. Ann Math statist, 1971,(42):1328～1338

[39] Jurečková J. Asymptotic relations of M-estimates and R-estimates in linear regression model. Ann Statist ,1977,(5):464～472

[40] Kiefer J. On Bahadur's representation of sample quantiles. Ann Math Statist, 1967,(38):1323～1342

[41] Koenker R. A comparison of asymptotic testing methods for L_1-regression. In Stafistical Data Analysis Based on the L_1-Norm and Related Methods, North-Holland,1987,287～295

[42] Koenker R W and Bassett G W. Test of linear hypothesis and L_1-estimation, Econometrika,1982,(50):1577～1583

[43] Koenker R amd Portnoy S. M-estimation of multivariate regressions. J Amer Statist Assoc,1990,(85):1060～1068

[44] Lai T L, Robbins H and Wei C Z. Strong consistency of least squares estimates in multiple regression Ⅰ. J Multivariate Anal, 1979,(9):343～362

[45] Loève M. Probability Theory. 4-th ed. ,Springer-Verlag,1977

[46] Marrona R A. Yohai V J. Asymptotic behavior for regression and scale with random carrics. Z Wahrsch Verb Gebiete,1981,(56):7～20

[47] Mckean J W and Schrader R M. Least absolute errors analysis of variance. In Statistical Data Analysis Based on the L_1-Norm and Related Methods, North—Holland,1987,297～305

[48] Niemiro W. Asymptotics for M-estimators defined by convex minimization, Ann Statist,1992,(20):1514～1533

[49] Petrov V V. Sums of Independent Random Variables. Springer-verlag, 1975

[50] Pollard D. Asymptotics for least absolute deviation regression estimatous. Econom Theory,1991,(7):186～199

[51] Portnoy S. Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large . I. Consistency. Ann Statist ,1984,(12):1298~1309

[52] Portnoy S. Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large . Ⅱ. Normal Approximation. Ann Statist,1985,(13):1403~1417

[53] Rao C R. Tests of significance in multivariate analysis,Biometrika,1948,(35):58~79

[54] Rao C R. Linear Statistical Inference and Its Applications. 2nd ed , John Wiley, 1973

[55] Rao C R and Zhao L C. Approximation to the distribution of M-estimates in linear models by randomly weighted bootstrap. Sankhyā Ser A,1992,(54):323~331

[56] Rao C R and Zhao L C. On the consistency of M-estimates in a linear model obtained through an estimating equation. Statist Probab Letters, 1992,(14):79~84

[57] Rao C R and Zhao L C. Linear representation of M-estimates in linear models. Canadian J Statist, 1992,(20):359~368

[58] Relles D A. Robust regression by modified least squares. Ph D Thesis, Yale Univ, 1968

[59] Rockafellar R T. Convex Analysis. Princeton Univ Press,1970

[60] Schrader R M and Hettmansperger T P. Robust analysis of variance based upon a likelihood ratio criterion. Biometrika, 1980,(67):93~101

[61] Sen P K. On M tests in linear models. Biometrika, 1982,(69):245~248

[62] Singer J M and Sen P K. M-methods in multivariate linear models. J Multivariate Analysis, 1985,(17):168~184

[63] Stout W F. Almost Sure Convergence. Academic Press, 1974

[64] Welsh A H. On M-Processes and M-estimation. Ann Statist, 1989, (17): 337 361;Correction. Ann Statist, 1990,(18):1500

[65] Wu Y. Strong consistency and exponential rate of the "minimum L_1-norm" estimates in linear regression models. Comput Statist Data Anal,1988,(6):285~295

[66] Yohai V J. Robust (M) estimates for the general linear model. Notas de Metemática,No. 23, Univ Nacional de La Plata,Facultad de Ciencias Exactas y Naturales,1972

[67] Yohai V J and Maronna R A. Asymptotic behavior of M-estimators for the linear model,Ann Statist,1979,(7):258~268

[68] Zhao L C aod Chen X R. Asymptotic behavior of M-test statistics in linear models. J Combin. Inform Syst Sciences,1991,(16):234~248

[69] Zhao L C, Rao C R and Chen X R. A note on the consistency of M-estimates in linear models. In Stochastic Process, A Festschrift in Honour of Gopinath Kallianpur, Springer-Verlag, 1993, 359～367